OXFORD

SECOND EDITION

COGNITIVE PSYCHOLOGY

EDITED BY

NICK BRAISBY & ANGUS GELLATLY

SHOUT
SHARE
IT

COGNITIVE PSYCHOLOGY

This page intentionally left blank

COGNITIVE PSYCHOLOGY

SECOND EDITION

EDITED BY

NICK BRAISBY AND ANGUS GELLATLY

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© Oxford University Press 2012

The moral rights of the authors have been asserted

First Edition copyright 2005
Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law, by licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloguing in Publication Data
Library of Congress Control Number: 2011943537

ISBN 978–0–19–923699–2

Printed in Italy by
L.E.G.O. S.p.A.—Lavis TN

This edition is adapted and reproduced from *Cognitive Psychology* edited by Nick Braisby and
Angus Gellatly (2005) and *Cognitive Psychology: a Methods Companion* edited by Nick Braisby (2005),
by permission of the Open University.

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

PREFACE TO THE SECOND EDITION

This is a new edition of a book originally published by Oxford University Press in conjunction with the Open University to serve as the core text for a Level 3 Open University course in cognitive psychology. The new edition is published solely by Oxford University Press and, while retaining what we consider to be the best features of the original, differs considerably from its predecessor. There are a number of aspects of the design of this text that we hope will serve students learning about cognitive psychology and educators teaching the subject.

Book structure

The chapters in this book are organized in six parts. The first four parts focus on broad and well-established topic areas within cognitive psychology, such as perceptual processes, language, memory, and thinking and reasoning. All these chapters have been updated, and in some cases considerably revised, since the first edition. The fifth part is entirely new. It comprises three chapters reflecting the increasing and exciting interplay between cognitive psychology and the cognitive neurosciences. The sixth part comprises five updated and revised chapters that consider a range of challenges, themes, and issues – topics that have been thought to present challenges to the cognitive approach, such as emotion and consciousness; themes such as cognitive modelling and modularity; and issues such as the relation of cognition to biology.

The first chapter is not located in one of these parts. It attempts to give a historical and conceptual introduction to cognitive psychology, laying out the foundations of the subject and raising some of the important themes and issues that are revisited in later chapters. Some of these themes are also developed in the introductions to each of the subsequent parts; we recommend that students read these introductions prior to reading their associated parts, and re-read them afterwards.

Chapter structure

Each chapter has been structured according to certain conventions.

An **emboldened term** signifies the introduction of a key concept or term that is either explicitly or implicitly defined in the surrounding text and/or explained in the glossary at the end of the book. The locations of these defined terms are also flagged in bold in the index.

Each chapter contains a number of **activities**. Often these may be simple thought exercises that may take no more than a minute or so. Others are more involved. Each activity has been integrated into the design of the chapter and is aimed at enhancing students' understanding of the material. We recommend that student readers attempt as many of these activities as possible and, where appropriate, revisit them after completing each chapter. Educators may wish to select these activities as the basis for in-class exercises.

The chapters in this book also make use of **text boxes**. Each box has been written to amplify a particular aspect of the material without interrupting the ongoing narrative. Though the boxes illuminate a wide range of issues, many focus on aspects of **research studies** and **methods**. Students may find that they wish to finish a section before reading a particular box.

Each substantive main section finishes with a **section summary**, often a bullet point list reminding the student of the key points established in that section. We hope that students will use these as useful barometers of their understanding and re-read sections where the summary points are not clearly understood.

Each chapter makes a number of explicit **links** to other chapters in the book, often to specific numbered sections. It would be tedious in the extreme to continually follow each and every link, flicking to the relevant pages and reading the relevant 'linked' section. Rather, these links are intended to help students perceive the interconnected nature of cognitive psychology, identifying connections between topics that otherwise might seem disparate. Of course, we hope that students will be motivated to follow some of these links, either on first reading or on a later reading, perhaps as a revision aid.

As well as a list of references, each chapter ends with some specific suggestions for **further reading**. While each chapter is designed to be self-contained, inevitably some issues get less attention than they deserve, and so interested readers may wish to pursue some of these suggestions for a more in-depth treatment. Moreover, it is always worth approaching a topic from more than one direction – consulting different texts, including other general texts on cognitive psychology, can help achieve a richer understanding and we recommend this approach to all students. Educators may wish to use these further readings as the basis for seminar discussions.

Supporting a course in cognitive psychology

There are few restrictions on how one might use this text to support the teaching of a course in cognitive psychology. The chapters in this book may be tackled in a number of different orders. Depending on the focus of the course, particular parts may be omitted, or particular chapters omitted from a given part or parts. The book as a whole presupposes relatively little prior knowledge of cognitive psychology on the part of the student. However, in some instances, later chapters may presuppose some limited knowledge of related earlier chapters, though this is usually explicitly indicated. Similarly, while all chapters are designed to be taught at the same level, later chapters may tackle issues considered too complex for the earlier chapters. By focusing more on earlier or later chapters, courses can vary somewhat in the degree of difficulty of the material they present.

Online Resource Centre

Associated with the book is an Online Resource Centre that contains much additional material (www.oxfordtextbooks.co.uk/orc/braisby2e/). This can be used to further students' understanding and may be used in presenting a course in cognitive psychology. Student resources include a hyperlinked bibliography and flashcard glossary, and lecturer resources include figures from the book available to download and a test bank of multiple-choice questions.

New to this edition

All of the chapters in this book have either been updated since the first edition or are entirely new to this edition. Chapters 1 to 12 have all been updated with additional material reporting new work in the area and provide the relevant recent references. Many of the chapters also include entirely new sections. Chapter 2 now has a section linking attention, consciousness, and hypnosis. Chapter 3 contains a new section describing the relevance of research on perception for aircraft safety. The topics covered in Chapter 4 have been updated and new material added on the recognition of familiar and unfamiliar faces. Chapter 5 incorporates a new section on exemplar theories, and significant new material relating to essentialism and to individual differences in categorization. The material covered in Chapter 6 has been updated and new information added on the representations involved in spoken word recognition. Chapter 7 has a new section on future directions in research into how language is used in conversation. Chapter 8 has been extensively revised and includes new sections on self-referential encoding, comparison and evaluation of different accounts of memory systems, and an analysis of signal detection accounts of memory performance. A new section has been added to Chapter 9 describing a recently added component of the working memory model, the episodic buffer. In Chapter 10, new sections have been added on analogical reasoning and theories of how insight operates in problem solving. Chapter 11 contains fresh material on prospect theory and a considerably extended discussion of the overconfidence effect in decision making, while Chapter 12 has a similarly expanded discussion of integrative approaches to reasoning.

Section 5 contains two entirely new chapters (13 and 15) that have been specially commissioned for this edition of *Cognitive Psychology*. In addition, Chapter 14 – a much revised version of a chapter originally published as part of a companion volume to the first edition – provides an updated account of recent developments in neuroimaging techniques, including near infrared spectroscopy, a technique particularly suited for use with infants and younger children.

The chapters in Section 6 have been revised in a variety of ways. Chapter 16 incorporates new material on cognitive biases, mood congruent memory, and emotion regulation. Chapter 17 has been updated throughout and now includes discussion of life scripts theory, as well as neuropsychological evidence concerning autobiographical memory. Chapter 18 now contains a discussion of recent work on techniques to detect consciousness, including the case of vegetative state patients. Chapter 19 has been revised to reflect recent changes in the ACT-R architecture, as well as its application to the development of mobile phone texting skills. Finally, Chapter 20 has been revised following the untimely death of the first author, Tony Stone. Under Martin Davies's authorship, new material has been added to provide extended treatments of connectionism, modularity, and cognitive neuropsychology.

Acknowledgements

There are a very considerable number of people we would like to thank, without whose contributions this second edition would not have been possible. We would like to thank all of those involved in the production of the first edition and of course all of the authors, some of whom are new to this second edition, for their work in updating and revising their chapters. We would like to say a special thank you to the hard-working team at Oxford University Press: Holly, Laura, Bethan, Siân, and Jonathan. We cannot recall how many times this project, and especially the editors, have been kept on track by their hard work behind the scenes, and their gentle but insistent prompting.

Lastly, we would like to pay a special tribute to the late Tony Stone, first author of Chapter 20. Tony was a fine illustration of the permeability of cognitive psychology; its openness to contributions from other disciplines. Tony was first and foremost a philosopher, but one who had increasingly turned his mind to the subject matter of cognitive psychology. His chapter on theoretical issues in the first edition of this book was very highly regarded and perfectly demonstrated his skills of analysis and communication. Tony was excited by cognitive psychology in part because it wrestles with deep questions concerning the nature of the mind, and we are sure he would be delighted if this volume served to stimulate a similar interest in the next generation of scholars. We hope it will.

Nick Braisby and Angus Gellatly, July 2011

PREFACE TO THE FIRST EDITION

This book has been produced as the core text for the Open University's level 3 course in Cognitive Psychology (DD303). However, it has been designed to serve students taking other courses in cognitive psychology as well, either as essential or recommended reading. There are a number of features of the design of this text that we hope will serve students learning about cognitive psychology and educators teaching the subject.

Book structure

The chapters in this book are organized in five parts. The first four parts focus on broad and well-established topic areas within cognitive psychology, such as perceptual processes and memory. The fifth considers a range of challenges, themes, and issues – topics that have been thought to present challenges to the cognitive approach, such as emotion and consciousness; themes such as cognitive modelling and modularity; and issues such as the relation of cognition to biology.

The first chapter is not located in one of these parts. It attempts to give a historical and conceptual introduction to cognitive psychology, laying out the foundations of the subject and raising some of the important themes and issues that are revisited in later chapters. Some of these themes are also developed in the introductions to each of the subsequent parts; we recommend that students read these introductions prior to reading their associated parts and re-read them afterwards.

Chapter structure

Each chapter has been structured according to certain conventions.

An emboldened term signifies the introduction of a key concept or term that is either explicitly or implicitly defined in the surrounding text. The locations of these defined terms are also flagged in bold in the index.

Each chapter contains a number of activities. Often these may be simple thought exercises that may take no more than a minute or so. Others are more involved. Each activity has been integrated into the design of the chapter and is aimed at enhancing students' understanding of the material. We recommend that student readers attempt as many of these activities as possible and, where appropriate, revisit them after completing each chapter.

The chapters in this book also make use of text boxes. Each box has been written to amplify a particular aspect of the material without interrupting the ongoing narrative. Though the boxes illuminate a wide range of issues, many focus on aspects of research studies and methods. Students may find that they wish to finish a section before reading a particular box.

Each substantive main section finishes with a section summary, often a bullet point list reminding the student of the key points established in that section. We hope that students will use these as useful barometers of their understanding and re-read sections where the summary points are not clearly understood.

Each chapter makes a number of explicit links to other chapters in the book, often to specific numbered sections. It would be tedious in the extreme to continually follow each and every link, flicking to the relevant pages and reading the relevant 'linked' section. Rather, these links are intended to help students perceive the interconnected nature of cognitive psychology, identifying connections between topics that otherwise might seem disparate. Of course, we hope that students will be motivated to follow some of these links, either on first reading or on a later reading, perhaps as a revision aid.

As well as a list of references, each chapter ends with some specific suggestions for further reading. While each chapter is designed to be self-contained, inevitably some issues get less attention than they deserve, and so interested readers may wish to pursue some of these suggestions for a more in-depth treatment. Moreover, it is always worth approaching a topic from more than one direction – consulting different texts, including other general texts on cognitive psychology, can help achieve a richer understanding and we recommend this approach to all students.

Supporting a course in cognitive psychology

There are few restrictions on how one might use this text to support the teaching of a course in cognitive psychology. The chapters in this book may be tackled in a number of different orders. Depending on the focus of the course, particular parts may be omitted, or particular chapters omitted from a given part or parts. The book as a whole presupposes relatively little prior knowledge of cognitive psychology on the part of the student. However, in some instances, later chapters may presuppose some limited knowledge of related earlier chapters, though this is usually explicitly indicated. Similarly, while all chapters are designed to be taught at the same level, later chapters may tackle issues considered too complex for the earlier chapters. By focusing more on earlier or later chapters, courses can vary somewhat in the degree of difficulty of the material they present.

Companion volume

Accompanying this book is a companion publication, *Cognitive Psychology: A Methods Companion*, also published by Oxford University Press and also designed as a key teaching text for the Open University's level 3 course in Cognitive Psychology. The *Methods Companion* considers in detail a number of key methodological issues in cognitive psychology, including ethics, connectionism, symbolic modelling, neuroimaging, neuropsychology, and statistics.

Companion web site

This book and the *Methods Companion* are associated with a companion web site that contains much additional material. This can be used to further students' understanding and may be used in presenting a course in cognitive psychology. Materials include electronic versions of figures, experiment and data files, and software for running cognitive models.

Acknowledgements

Developing the Open University's level 3 course in Cognitive Psychology (DD303) has been a major undertaking, involving the production of two books, various pieces of software and associated files, audio materials, web sites and web-based materials, and numerous other additional items and activities. To say that such a course, and that this text, could not have been produced without the help and cooperation of a large number of people is an understatement. We extend our grateful and sincere thanks to all those who have made this enterprise possible.

Nick Braisby and Angus Gellatly, Milton Keynes, January 2005

This page intentionally left blank

CONTENTS IN BRIEF

List of contributors xxvii

1 FOUNDATIONS OF COGNITIVE PSYCHOLOGY *Nick Braisby and Angus Gellatly* 1

PART 1 PERCEPTUAL PROCESSES

INTRODUCTION 30

2 ATTENTION *Peter Naish* 33

3 PERCEPTION *Graham Pike, Graham Edgar, and Helen Edgar* 65

4 RECOGNITION *Graham Pike and Nicola Brace* 100

PART 2 CONCEPTS AND LANGUAGE

INTRODUCTION 138

5 CONCEPTS *Nick Braisby* 141

6 LANGUAGE PROCESSING *Gareth Gaskell and Helen Brown* 169

7 LANGUAGE IN ACTION *Simon Garrod and Tony Sanford* 197

PART 3 MEMORY

INTRODUCTION 226

8 LONG-TERM MEMORY Encoding to retrieval *Andrew Rutherford, Gerasimos Markopoulos, Davide Bruno, and Mirjam Brady-Van den Bos* 229

9 WORKING MEMORY *Graham J. Hitch* 266

PART 4 THINKING

INTRODUCTION 298

10 PROBLEM SOLVING *Alison J.K. Green and Ken Gilhooly* 301

11 JUDGEMENT AND DECISION MAKING *Peter Ayton* 334

12 REASONING *Mike Oaksford* 366

PART 5 CONTRIBUTIONS FROM THE COGNITIVE NEUROSCIENCES

INTRODUCTION 394

13 COGNITIVE NEUROPSYCHOLOGY *Gianna Cocchini* 399

14 NEUROIMAGING *Ingrid S. Johnsrude and Olaf Hauk* 430

15 COGNITIVE NEUROPSYCHIATRY *David Okai, Angus Gellatly, and Anthony David* 468

PART 6 CHALLENGES, THEMES, AND ISSUES

INTRODUCTION 504

16 COGNITION AND EMOTION *Jenny Yiend, Bundy Mackintosh, and George Savulich* 507

17 AUTOBIOGRAPHICAL MEMORY AND THE SELF *Clare J. Rathbone, Chris J.A. Moulin, Martin A. Conway, and Emily A. Holmes* 546

18 CONSCIOUSNESS *Jackie Andrade* 577

19 COGNITIVE MODELLING AND COGNITIVE ARCHITECTURES *Paul Mulholland and Stuart Watt* 606

20 THEORETICAL ISSUES IN COGNITIVE PSYCHOLOGY *Tony Stone and Martin Davies* 639

Epilogue 681

Glossary 685

Index 707

CONTENTS IN FULL

List of contributors xxvii

1 FOUNDATIONS OF COGNITIVE PSYCHOLOGY
Nick Braisby and Angus Gellatly 1

1 Introduction 1
2 What is cognitive psychology? 2
3 A brief history of cognitive psychology 8
3.1 Introspectionism 8
3.2 Gestalt psychology 9
3.3 Behaviourism 10
3.4 The return of the cognitive 12
4 Science, models, and the mind 17
5 The cognitive approach 20
5.1 Representation 20
5.2 Computation 21
6 Level-dependent explanations 23
6.1 The computational level 23
6.2 The algorithmic level 24
6.3 The implementational level 25
6.4 Using Marr's levels 25
7 Conclusions 26
FURTHER READING 27
REFERENCES 27

PART 1 PERCEPTUAL PROCESSES

INTRODUCTION 30

2 ATTENTION *Peter Naish* 33
1 Auditory attention 33
1.1 Disentangling sounds 34
1.2 Attending to sounds 36
1.3 Eavesdropping on the unattended message 38
2 Visual attention 40
2.1 Knowing about unseen information 41
2.2 Towards a theory of parallel processing 43
2.3 Rapid serial visual presentation 44
2.4 Masking and attention 47

3 Integrating information in clearly-seen displays 48
  3.1 Serial and parallel search 48
  3.2 Non-target effects 50
  3.3 The ‘flanker’ effect 50
4 Attention and distraction 53
  4.1 The effects of irrelevant speech 53
  4.2 Attending across modalities 54
5 The neurology of attention 55
  5.1 The effects of brain damage 56
  5.2 Event-related potentials 58
6 Drawing it all together 58
  6.1 Attention as the generator of consciousness 60
  6.2 Conclusions 61
FURTHER READING 62
REFERENCES 62

3 PERCEPTION *Graham Pike, Graham Edgar, and Helen Edgar* 65
1 Introduction 65
  1.1 Perceiving and sensing 67
  1.2 The eye 68
  1.3 Approaches to perception 69
2 The Gestalt approach to perception 71
3 Gibson’s theory of perception 73
  3.1 An ecological approach 74
  3.2 The optic array and invariant information 74
  3.3 Flow in the ambient optic array 78
  3.4 Affordances and resonance 80
4 Marr’s theory of perception 81
  4.1 The grey level description 82
  4.2 The primal sketch 82
  4.3 The 2½D sketch 85
  4.4 Evaluating Marr’s approach 86
5 Constructivist approaches to perception 87
6 The physiology of the human visual system 90
  6.1 From the eye to the brain 90
  6.2 The dorsal and ventral streams 91
  6.3 The relationship between visual pathways and theories of perception 92
  6.4 A dual-process approach? 92
  6.5 Applying perceptual research: a case study of an ‘aircraft proximity event’ 93
  6.6 Combining bottom-up and top-down processing 94

7 Conclusion 96

FURTHER READING 97

REFERENCES 98

4 RECOGNITION *Graham Pike and Nicola Brace* 100

1 Introduction 100

1.1 Recognition in the wider context of cognition 101

2 Different types of recognition 102

2.1 Object and face recognition 102

2.2 Active processing – recognizing objects by touch 104

2.3 Recognizing two-dimensional objects 105

2.4 Object-centred vs viewer-centred descriptions 107

3 Recognizing three-dimensional objects 108

3.1 Marr and Nishihara's theory 109

3.2 Evaluating Marr and Nishihara's theory 114

3.3 Biederman's theory 114

3.4 Evaluating Biederman's theory and the object-centred approach 116

4 Face recognition 118

4.1 Comparing object and face recognition 118

4.2 Recognizing familiar and unfamiliar faces 119

5 Modelling face recognition 121

5.1 Young *et al.*'s (1985) diary study 121

5.2 A cognitive model of face recognition 122

5.3 A connectionist model of face recognition 124

6 Neuropsychological evidence 126

6.1 Separate routes for identification and emotional expression 126

6.2 Covert face recognition 127

6.3 Evaluating the IAC model 129

7 Are faces 'special'? 129

8 Conclusion 133

FURTHER READING 134

REFERENCES 134

PART 2 CONCEPTS AND LANGUAGE

INTRODUCTION 138

5 CONCEPTS *Nick Braisby* 141

1 Introduction 141

1.1 Concepts, categories, and words 141

1.2 Categorization 142

1.3 The wider story of concepts 144

1.4 Concepts and cognition 144

2 Explaining categorization 146
    2.1 Similarity I: the classical view of concepts 146
    2.2 Similarity II: prototype theories of concepts 151
    2.3 Similarity III: exemplar theories 155
    2.4 Commonsense theories 155
    2.5 Psychological essentialism 158

3 Where next? 162
    3.1 Is all categorization the same? 162
    3.2 Are all concepts the same? 163
    3.3 Are all categorizers the same? 164

4 Conclusion 165

FURTHER READING 166

REFERENCES 166

6 LANGUAGE PROCESSING *Gareth Gaskell and Helen Brown* 169

1 Introduction 169

2 Word recognition 170
    2.1 Spoken word recognition 170
    2.2 Visual word recognition 176

3 The mental lexicon 182
    3.1 Morphology 182
    3.2 Accessing word meanings 184

4 Sentence comprehension 187
    4.1 Syntax 188
    4.2 Models of parsing 189
    4.3 Is parsing autonomous? 191
    4.4 Constraints on parsing 192

5 Conclusion 193

FURTHER READING 194

REFERENCES 194

7 LANGUAGE IN ACTION *Simon Garrod and Tony Sanford* 197

1 Introduction 197

2 Written language and discourse 198
    2.1 Processes underlying text interpretation 199
    2.2 Special topics 204

3 Language production and dialogue 208
    3.1 Language production as a self-contained process 208
    3.2 The challenge of dialogue 214
    3.3 The monologue dialogue distinction and group decision making 218

4 Conclusions 220

FURTHER READING 221

REFERENCES 221

PART 3 MEMORY

INTRODUCTION 226

8 LONG-TERM MEMORY
Encoding to retrieval
Andrew Rutherford, Gerasimos Markopoulos, Davide Bruno, and Mirjam Brady-Van den Bos 229

- 1 Introduction 229
- 2 Encoding 230
  - 2.1 Levels of processing 231
  - 2.2 Item-specific and relational processing 233
  - 2.3 Other forms of encoding processing 234
- 3 Memory stores and systems 235
  - 3.1 Declarative and non-declarative memory systems 236
  - 3.2 Multiple memory systems 238
  - 3.3 Comparing declarative and non-declarative, and multiple memory systems 238
- 4 Retrieval 243
  - 4.1 Encoding specificity and transfer appropriate processing 243
  - 4.2 The diagnostic value of retrieval cues 245
  - 4.3 Accounts of recognition memory 246
  - 4.4 Remembering and knowing 250
- 5 Implicit memory 252
  - 5.1 Perceptual and conceptual incidental tests 253
  - 5.2 Intentionality and awareness 254
  - 5.3 Accounts of implicit memory 255
- 6 Conclusions 258

FURTHER READING 259

REFERENCES 259

9 WORKING MEMORY *Graham J. Hitch* 266

- 1 Introduction 266
  - 1.1 Human memory as a multifaceted system 266
  - 1.2 Distinction between short-term and long-term memory 267
  - 1.3 Working memory as more than STM 269
- 2 The structure of working memory 272
  - 2.1 A multi-component model 272
  - 2.2 Phonological working memory 274
  - 2.3 Executive processes 279

3 Vocabulary acquisition 285
- 3.1 Neuropsychological evidence 286
- 3.2 Individual differences 286
- 3.3 Experimental studies 286

4 Modelling the phonological loop 287
- 4.1 Serial order 288

5 Conclusion 290

FURTHER READING 291

REFERENCES 291

PART 4 THINKING

INTRODUCTION 298

10 PROBLEM SOLVING *Alison J.K. Green and Ken Gilhooly* 301

1 Introduction 301
- 1.1 What is a ‘problem’? 302
- 1.2 Protocol analysis in problem-solving research 304

2 ‘Simple’ problem solving 306
- 2.1 The Gestalt legacy 306
- 2.2 Representation in puzzle problem solving 309
- 2.3 The information processing approach: problem solving as search 310
- 2.4 Information processing approaches to insight 311
- 2.5 Insight: inappropriate heuristics or misleading representations? 313
- 2.6 Testing two insight theories 315

3 Analogical problem solving 316
- 3.1 Analogies in problem solving 317
- 3.2 How do analogies work? 318

4 ‘Complex’ problem solving 319
- 4.1 The role of knowledge in expert problem solving 320

5 Prospects for problem-solving research 324
- 5.1 Limitations of the modal model 324
- 5.2 Individual differences 326

6 Conclusions 329

FURTHER READING 330

REFERENCES 330

11 JUDGEMENT AND DECISION MAKING *Peter Ayton* 334

1 Introduction 334
- 1.1 Theories of decision making 335
- 1.2 Supporting decision making 335

2 Normative theory of choice under risk 336

2.1 Prescriptive application of normative theory: decision analysis 336

2.2 Axioms underlying subjective expected utility theory 339

2.3 Violations of the axioms 339

3 Findings from behavioural decision research 342

3.1 The ‘preference reversal phenomenon’ 343

3.2 Causes of anomalies in choice 343

4 Prospect theory 346

4.1 Prospect theory and ‘loss aversion’ 346

4.2 Prospect theory and probability weighting 347

4.3 ‘Framing’ effects 348

4.4 Prospect theory and experience-based decisions 348

5 Judgement under uncertainty 349

5.1 Judging probabilities and Bayes’ Theorem 349

5.2 Does Bayes’ Theorem describe human judgement? 351

5.3 Heuristics and biases 352

5.4 Evaluating the heuristics and biases account 354

5.5 Overconfidence 356

6 Fast and frugal theories of decision making 360

7 Conclusion 361

FURTHER READING 362

REFERENCES 362

12 REASONING *Mike Oaksford* 366

1 Introduction 366

1.1 Reasoning in everyday life 367

2 Deductive reasoning and logic 368

3 Psychological theories of reasoning 370

3.1 Mental logic 371

3.2 Mental models 371

3.3 The probabilistic approach 371

4 Conditional inference 372

4.1 The abstract conditional inference task 372

4.2 Everyday reasoning and the suppression effect 375

5 Wason’s selection task 379

5.1 The abstract selection task 379

5.2 The deontic selection task 382

6 Discussion 387

6.1 Theoretical evaluation 387

6.2 Integrative approaches 389

FURTHER READING 390

REFERENCES 391

PART 5 CONTRIBUTIONS FROM THE COGNITIVE NEUROSCIENCES

INTRODUCTION 394

13 COGNITIVE NEUROPSYCHOLOGY *Gianna Cocchini* 399

1 Introduction: the lesion method and cognitive neuropsychology 399
2 Historical milestones 404
3 Assumptions in cognitive neuropsychology 407
3.1 Modularity 407
3.2 Subtractivity and transparency 408
3.3 Anatomical and functional syndromes 410
4 Methods in cognitive neuropsychology 414
4.1 Association, dissociation, and double dissociation 414
4.2 Group studies versus single-case studies 417
5 Further crucial issues for cognitive neuropsychology 420
5.1 Compensatory strategies and related issues 420
5.2 Pre-morbid pathological conditions and previous brain damage 421
5.3 Combination of deficits: the contribution and interaction of different deficits 422
6 Measuring behavioural responses 423
7 Conclusions 425

FURTHER READING 426
REFERENCES 426

14 NEUROIMAGING *Ingrid S. Johnsrude and Olaf Hauk* 430

1 Introduction 430
2 EEG and MEG: measuring the timing of electrical activity in the brain 433
2.1 Collecting data 433
2.2 Basic analysis of EEG and MEG data 437
2.3 Estimating neuronal sources: the inverse problem 440
2.4 Statistical analysis of EEG/MEG data 444
2.5 The auditory oddball paradigm 444
3 Techniques based on metabolism and blood supply 447
3.1 Experimental design 447
3.2 fMRI: functional magnetic resonance imaging 450
3.3 PET: positron emission tomography 458
3.4 SPECT: single photon emission computerized tomography 460
3.5 NIRS: near-infrared spectroscopy 460
4 Transcranial magnetic stimulation (TMS) 461

5 Choosing a technique 462
6 Summary and conclusions 464
FURTHER READING 465
REFERENCES 466

15 COGNITIVE NEUROPSYCHIATRY *David Okai, Angus Gellatly, and Anthony David* 468

1 Introduction 468
2 Focal and diffuse brain injury 470
  2.1 Frontal lobe 470
  2.2 Parietal lobe 471
  2.3 Temporal lobes 472
  2.4 Occipital lobe 472
  2.5 Brain stem and cerebellum 472
3 Dementia 473
  3.1 Alzheimer’s disease 474
  3.2 Vascular dementia 475
  3.3 Dementia with Lewy bodies (DLB) 475
  3.4 Fronto-temporal dementia 476
  3.5 Mild cognitive impairment 476
4 Psychotic illnesses 477
  4.1 Normal condition 478
  4.2 Hallucinations 479
  4.3 Formal thought disorder 480
  4.4 Negative symptoms 480
  4.5 Delusions 481
  4.6 Monothematic delusions: a two-factor approach 482
5 Depression 484
  5.1 Emotion 484
  5.2 Neuroanatomical correlates of depression 485
  5.3 Depression phenotype and how it relates to cognition 486
  5.4 Stroke 486
  5.5 Traumatic brain injury 488
  5.6 Pseudodementia 490
6 Disinhibitory psychopathologies 491
  6.1 Obsessive–compulsive disorder 491
  6.2 Hoarding 493
  6.3 Impulse control disorders 493
7 Conversion disorder 496
8 Summary 497
FURTHER READING 497
REFERENCES 497

PART 6 CHALLENGES, THEMES, AND ISSUES

INTRODUCTION 504

16 COGNITION AND EMOTION *Jenny Yiend, Bundy Mackintosh, and George Savulich* 507

1 Introduction 507
- 1.1 Components of emotion 508

2 Different emotions 512
- 2.1 Basic emotions 512
- 2.2 Verbal labels 517
- 2.3 The dimensional approach 518

3 The function of emotions 520
- 3.1 Emotions alter goals 520
- 3.2 Emotions mobilize physiological resources 521
- 3.3 Emotional expressions as communication 521
- 3.4 Emotions as information 522
- 3.5 What is the function of emotional feelings? 523

4 Emotion influences cognition 524
- 4.1 Some important concepts 524
- 4.2 Memory 526
- 4.3 Attention 530
- 4.4 Semantic interpretation 532

5 Does cognition influence emotion? 535
- 5.1 A look at some historical answers 535
- 5.2 A clash of minds: the cognition/emotion debate 539

6 General summary 542

FURTHER READING 543

REFERENCES 543

17 AUTOBIOGRAPHICAL MEMORY AND THE SELF *Clare J. Rathbone, Chris J.A. Moulin, Martin A. Conway, and Emily A. Holmes* 546

1 What are autobiographical memories? 546

2 Autobiographical memory across the lifespan 548
- 2.1 Childhood amnesia 550
- 2.2 The reminiscence bump 550
- 2.3 Recency 552

3 Autobiographical knowledge, episodic memory, the working self, and memory construction 552
- 3.1 Autobiographical knowledge 554
- 3.2 Episodic and semantic memory 556
- 3.3 The working self 558
- 3.4 Constructing autobiographical memories 560

4 Autobiographical memory in distress 564
- 4.1 Traumatic event 565
- 4.2 Response at the time of trauma 565
- 4.3 Subsequent psychological symptoms 566
- 4.4 Impact of symptoms 567
- 4.5 The nature of intrusive trauma memories 568

5 What next for autobiographical memory and the self? 570

FURTHER READING 572

REFERENCES 572

18 CONSCIOUSNESS *Jackie Andrade* 577

1 Introduction 577
- 1.1 Defining consciousness 578
- 1.2 Philosophical approaches to consciousness 580
- 1.3 The place of consciousness within cognitive psychology 581

2 Empirical research: cognitive studies of consciousness 583
- 2.1 Implicit cognition 583
- 2.2 Controlled versus automatic processing 589
- 2.3 The neuropsychology of consciousness 591

3 What is consciousness for? 593
- 3.1 Altered states of consciousness 597

4 Cognitive theories of consciousness 600

5 Conclusion: what can cognitive psychology tell us about consciousness? 603

FURTHER READING 603

REFERENCES 604

19 COGNITIVE MODELLING AND COGNITIVE ARCHITECTURES *Paul Mulholland and Stuart Watt* 606

1 What is cognitive modelling? 606
- 1.1 Parallel distributed processing 607
- 1.2 Rule-based systems 609
- 1.3 Cognitive architectures 610

2 An overview of ACT-R 611
- 2.1 A brief history of ACT-R 611
- 2.2 The architecture of ACT-R 611
- 2.3 Declarative memory 612
- 2.4 Procedural memory 613
- 2.5 Goals and the goal module 615
- 2.6 Perception and action 616

3 ACT-R accounts of memory phenomena 618
- 3.1 Declarative representation of lists 619
- 3.2 Production rules for the rehearsal and retrieval of lists 620

3.3 List activation 622
3.4 Running the model 623
3.5 Evaluation of the ACT-R approach to modelling memory 624

4 Learning and problem solving 625
4.1 Production compilation 625
4.2 An example of human problem-solving behaviour: addition by counting 627
4.3 Learning in an everyday context: the development of mobile phone texting skills 629
4.4 Models of learning and problem solving in practice 631

5 A comparison of ACT-R and PDP 632
6 When is a model a good model? 634
7 Conclusions 636

FURTHER READING 637
REFERENCES 637

20 THEORETICAL ISSUES IN COGNITIVE PSYCHOLOGY
Tony Stone and Martin Davies 639

1 Introduction 639
2 Computation and cognition 641
2.1 The computational model of the mind 642
2.2 Connectionist modelling 643
2.3 The past-tense debate: connectionism versus the CMM 644
2.4 Rules in connectionist networks 647
2.5 Connectionism, structure, and compositionality 649

3 Modularity 651
3.1 An outline of Fodor's theory of modularity 652
3.2 The central systems 656
3.3 Debates about modularity 659

4 Cognitive neuropsychology 663
4.1 Inference in cognitive neuropsychology 664
4.2 Cognitive neuropsychology and the past-tense debate 666

5 Cognitive psychology and the brain 669
5.1 Levels of explanation 669
5.2 Co-evolution and reduction 670
5.3 Is cognitive psychology just a historical staging post? 672
5.4 Is cognitive psychology autonomous from neurobiology? 673

6 Conclusion 674

FURTHER READING 675
REFERENCES 676

Epilogue 681
Glossary 685
Index 707

LIST OF CONTRIBUTORS

Professor Jackie Andrade
School of Psychology, Plymouth University

Professor Peter Ayton
Department of Psychology, City University

Dr Nicola Brace
Department of Psychology, The Open University

Dr Mirjam Brady-Van den Bos
School of Psychology, University of Aberdeen

Professor Nick Braisby
Department of Psychology, University of Winchester

Professor Helen Brown
Department of Psychology, University of York

Dr Davide Bruno
Nathan S. Kline Institute for Psychiatric Research; and Department of Psychiatry, New York University

Dr Gianna Cocchini
Psychology Department, Goldsmiths College, University of London

Professor Anthony David
Institute of Psychiatry, King's College, University of London

Professor Martin Davies
Faculty of Philosophy, University of Oxford

Dr Graham Edgar
Department of Natural and Social Science, University of Gloucestershire

Dr Helen Edgar
Department of Natural and Social Science, University of Gloucestershire

Professor Gareth Gaskell
Department of Psychology, University of York

Professor Simon Garrod
School of Psychology, University of Glasgow

Professor Angus Gellatly
Department of Psychology, Oxford Brookes University

Professor Ken Gilhooly
Department of Psychology, University of Hertfordshire

Dr Alison Green
Department of Psychology, The Open University

Dr Olaf Hauk
MRC Cognition and Brain Sciences Unit, Cambridge University

Professor Graham Hitch
Department of Psychology, University of York

Dr Ingrid S. Johnsrude
Department of Psychology, Queen's University, Ontario, Canada

Dr Bundy Mackintosh
School of Medicine, University of East Anglia

Dr Gerasimos Markopoulos
Department of Neurology, Otto-von-Guericke University

Dr Chris Moulin
Institute of Psychological Sciences, University of Leeds

Dr Paul Mulholland
Knowledge Media Institute and Centre for Research in Computing, The Open University

Dr Peter Naish
Sackler Centre for Consciousness Science, University of Sussex

Professor Mike Oaksford
Department of Psychological Sciences, Birkbeck College, University of London

Dr David Okai
Institute of Psychiatry, King's College, University of London

Dr Graham Pike
Department of Psychology, The Open University

Dr Clare Rathbone
School of Psychology and Clinical Language Sciences, University of Reading

Dr Andrew Rutherford
School of Psychology, Keele University

Professor Anthony J. Sanford
School of Psychology, University of Glasgow

Mr George Savulich
Institute of Psychiatry, King's College, University of London

Dr Stuart Watt
Ontario Institute for Cancer Research, Toronto, Ontario, Canada

Dr Jenny Yiend
Institute of Psychiatry, King's College, University of London

This page intentionally left blank

CHAPTER 1

FOUNDATIONS OF COGNITIVE PSYCHOLOGY

Nick Braisby and Angus Gellatly

1 INTRODUCTION

How does memory work? How do we understand language and produce it so that others can understand? How do we perceive our environment? How do we infer from patterns of light or sound the presence of objects in our environment, and their properties? How do we reason and solve problems? How do we think? And how might the study of abnormal functioning, of brain-damaged or psychiatric patients, shed light on these issues?

These are just some of the foundational questions that cognitive psychology examines. They are foundational partly because each concerns the nature of a basic psychological ability, abilities that we often take for granted, yet that are vital to our normal, healthy functioning and key to our understanding of what it means to be human. And they are foundational partly because they are important for psychology as a whole, and not just cognitive psychology. For instance, how can we hope to understand completely the behaviour of employees in an organization unless we first understand their perceptions and memories, and how they reason and attempt to solve problems? How can we understand the way in which people interact to shape one another's opinions if we do not understand how people understand and process language and how they make judgements?

Throughout this book, the various authors tackle these and other questions and show you how much of these foundations cognitive psychologists have so far uncovered. The book begins with an exploration of perceptual processes, moves to a discussion of categorization and language, through to memory, and then to thinking processes. It considers how cognitive psychology has been augmented by other methods of study, such as those used to examine patients suffering from brain damage or psychiatric conditions, and techniques that provide us with detailed images of brain activity and structure. Finally, the last part of the book is devoted to wider issues: to topics that have been thought to present a challenge to cognitive psychology – such as consciousness and emotion – and to some of the themes and theoretical questions that pervade the cognitive approach.

In this chapter we try to answer the question, 'What is cognitive psychology?'. In so doing, we outline some of the foundational assumptions that cognitive psychologists tend to make, as well as some of the reasons why it is such an important and fascinating subject – not least the fact that it raises many deep and important questions concerning the mind. We consider some of the issues that have attracted and continue to attract the interest of cognitive psychologists, and some of the assumptions they make in order to develop models and theories. We also reflect upon the cognitive approach in general and the kinds of explanation cognitive psychologists favour, briefly considering the relations between cognitive psychology and other sub-disciplines of psychology, and those between cognitive psychology and other disciplines (such as philosophy, computing, and linguistics).

There are many substantial issues that we only touch on – it is not easy to define the relationship between two academic disciplines, for example – and so we only hope to convey something of their flavour here. Our aim in this chapter is therefore merely to *introduce* cognitive psychology, to explain some of its key distinguishing features, and to uncover some of the many broad issues lying beneath its surface. You will obtain a richer and more complete overview of cognitive psychology from reading subsequent chapters, especially Chapter 20. You may find that the current chapter raises as many questions as it answers and that, as your reading of this book progresses, you periodically want to revisit this chapter to gain a better understanding of issues that, on first reading, seemed hazy. If this chapter only raises questions that you have in mind when you read subsequent chapters, and arouses your curiosity sufficiently that you periodically revisit it, it will have served its purpose well.

2 WHAT IS COGNITIVE PSYCHOLOGY?

What is cognitive psychology? Well, as with most questions, there can be short or long answers. The short, though not uncontentious, answer is that cognitive psychology is the branch of psychology devoted to the scientific study of the mind. Straightforward as this may seem, to understand the nature of cognitive psychology means digging deeper. And it is an excavation that raises all manner of substantial and

✱ ACTIVITY 1.1

Given the above definition that cognitive psychology is the scientific study of the mind, take a few minutes to write down some examples of what you would expect its characteristic features to be. For example, you might want to list what you take to be the characteristic features of a 'scientific' approach within psychology generally, and you might want to list some of the characteristic topics you would expect cognitive psychologists to study.

Keep your list ready to refer to as you read the rest of this chapter.

interesting issues – as diverse as the nature of normality and computation, and the importance of individual differences and brain images.

Activity 1.1 raises a number of interesting questions about the nature and scope of cognitive psychology. What does it mean for a psychology to be 'cognitive', for example? Did your list make any reference to normality? Well, when we say that cognitive psychology is the scientific study of the mind, this usually means 'normally-functioning human minds'. We can develop an understanding of the normal human mind in various ways: by studying people with normal minds and normal brains, but also by studying people with abnormal minds or abnormal brains, and by studying animals of other species, and even devices such as computers, with no brain at all. With respect to just this one issue – normality – cognitive psychology is clearly a broad enterprise. Box 1.1 gives a brief illustration of how evidence from people with brain damage can inform our understanding of normal cognition. Don't worry too much if you cannot follow all of the details at this stage – just try to get a feel for how cognitive psychologists have tried to relate evidence from brain-damaged patients to normal cognition.

'Cognitive psychology' can also be used to refer to activities in a variety of other disciplines and sub-disciplines (did your list refer to other disciplines?). Some sub-disciplines, like cognitive neuropsychology, developmental cognitive neuropsychology, cognitive neuropsychiatry, and cognitive neuroscience, include the cognitive signifier in their own titles. Others, such as behavioural neurobiology, linguistics, and artificial intelligence, do not; some practitioners of these might well object to finding themselves included under the cognitive psychology umbrella. As you will see in Chapter 5, uncertainty and negotiation regarding membership are characteristic of many if not all of our conceptual categories. Our advice is not to worry too much about such definitional issues at this stage, and perhaps not even later on. But one thing that is clear is that there is no easily identified boundary between cognitive psychology and work carried on in other disciplines with which cognitive psychologists frequently engage.

Your list of features of cognitive psychology may have referred to some of the methods that cognitive

BOX 1.1 RESEARCH STUDY Category-specific impairments I: neuropsychological methods

Warrington and Shallice (1984) describe four patients with specific impairments in recognizing living things. Because the impairment was thought to be specific to the category of living things, it has been called a **category-specific impairment**. One patient, JBR, experienced brain damage after suffering from herpes simplex encephalitis. As a result, when asked to name pictures, he correctly named only approximately 6 per cent of the pictures of living things, yet around 90 per cent of the pictures of non-living things. Other patients, though fewer of them, have been found to show an opposite impairment – that is, an impairment primarily to the category of non-living things (Hillis and Caramazza, 1991).

These studies have suggested to researchers that, in normal cognition, the categories of living and non-living things might be represented and/or processed differently. For example, one suggestion, that has since been much debated, was that in normal cognition the functional and sensory properties of categories are represented differently, and that living things tend to depend more on the sensory properties, while non-living things depend more on functional properties (Warrington and Shallice, 1984). This suggestion was also at first thought to help explain why JBR, on the assumption that he has an impairment for sensory properties, was also found to show impairments for some non-living categories, such as the categories of musical instruments and foods.

BOX 1.2 RESEARCH STUDY Category-specific impairments II: experimental and neuroimaging methods

Devlin *et al.* (2000) combined features of experimental and neuroimaging methods to investigate whether the categories of living and non-living things could be associated with representations in different parts of the brain. One technique they used was a lexical decision task. In this task, participants either hear or see strings of letters (e.g. they might see the strings 'warnd' or 'world') and have to judge whether each string is a word or not. Experimenters typically record both the judgement made and the amount of time participants take to make their response (perhaps by pressing the appropriate button on a keyboard or response pad). Another task, which Devlin *et al.* called a semantic categorization task, required participants, having seen three words presented one after another, to judge whether a fourth word belonged to the same category as the first three. Devlin *et al.* carefully matched words for word frequency and letter length. Whilst performing the lexical decision and semantic categorization tasks described above, participants were scanned using positron emission tomography (PET) technology. Another group of participants performed the semantic categorization task using pictures that were matched for visual complexity; these participants were scanned using functional magnetic resonance imaging (fMRI) technology. Both of these scanning technologies enable experimenters to identify regions of the brain that are particularly active during the performance of a task. Critically, Devlin *et al.* found no differences between the categories of living and non-living things in terms of active regions of the brain in either the PET study or the fMRI study (see Figures 1.1 and 1.2). So the differences in representation discussed in Box 1.1 may not be associated with different brain regions (or perhaps these techniques were not sensitive enough to detect such differences).

However, Kriegeskorte *et al.* (2008), examining brain activity both in human volunteers and monkeys, did find evidence that the brain represents categories in different ways. Focusing on the inferotemporal cortex, they found that in both species, viewing photographs of objects from the same category tended to elicit similar patterns of activity. Moreover, photographs of objects from different categories produced different patterns. Using a technique known as cluster analysis, Kriegeskorte *et al.* asked whether the patterns of activity themselves showed evidence of a category structure. The analysis showed that the patterns could be grouped according to a distinction similar to, but not the same as, that between living and non-living categories. Specifically, they found that both in humans and monkeys, patterns of activity in the inferotemporal cortex were different for animate

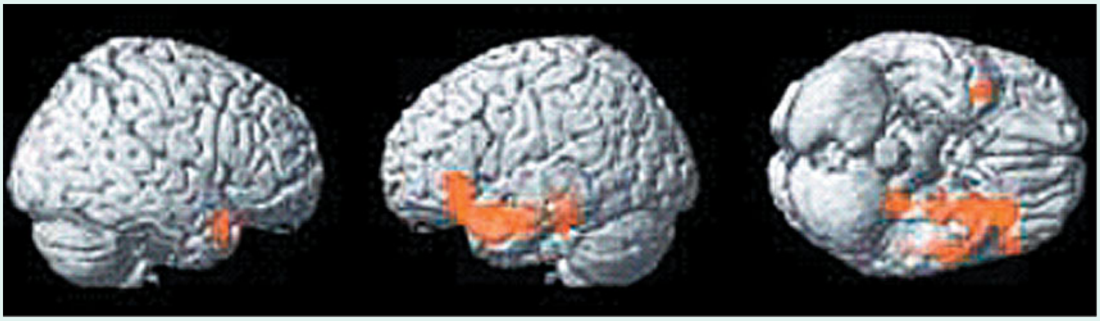


FIGURE 1.1 Results of PET studies using written words in lexical decision and semantic categorization tasks. Red areas show activated regions. No region showed a significant difference in activation between living and non-living things. *Source: Tyler and Moss, 2001, Figure 1*

(the categories of faces and bodies) and inanimate objects (though strangely, the 'inanimate' category derived from the brain activity of the human participants also included examples of non-human bodies, such as snakes and a wolf-like creature). See Chapter 14 for further discussion of this study.

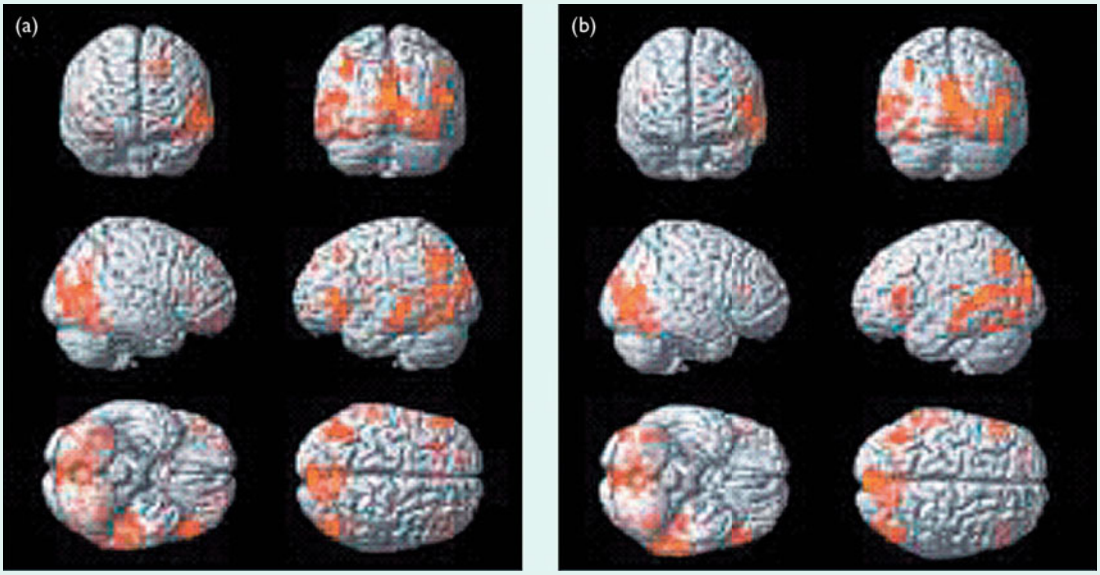


FIGURE 1.2 Results of fMRI study using pictures of living and non-living things in a semantic categorization task. The active brain regions are associated with non-living things (a) and living things (b). No regions showed significant differential activation for living and non-living things. *Source: Tyler and Moss, 2001, Figure 2*

psychologists employ: experiments, models (including computer models), neuropsychological investigations, and neuroimaging (or brain scans). Box 1.2 continues the discussion of category-specific impairments and describes a study that combines features of experimental and neuroimaging methods.

Box 1.3 describes a study employing cognitive modelling methods to examine category-specific impairments.

Boxes 1.1 to 1.3 illustrate some of the methods that will be referred to throughout this book, and about some of which we will say more later. But, perhaps more obvious than any of these issues, Activity 1.1 raises the question of the subject matter of cognitive psychology. What is it that cognitive psychologists study?

An easy way of answering this question (and one you might have adopted for Activity 1.1) is by scanning this book's table of contents. This will give you a good idea of the topics cognitive psychologists typically study, as, of course, will previous study of psychology. Certainly, the topics of perception, attention, language, categorization, reasoning, problem solving, and memory are central to the study of cognition. And cognition has broadened to include topics that have not always been seen as readily amenable to a cognitive approach (e.g. consciousness and emotion). The subsequent chapters will have much more to say about these issues than we can here. Activity 1.2 provides another way of thinking about the topics that interest cognitive psychologists.

Activity 1.2 shows how everyday behaviour can be explained in a number of different ways, involving many different kinds of cognitive process. In fact, all of the types of explanation referred to in the comment on Activity 1.2 are ones that will be developed at some length in this book. However, a corollary of the observations made in Activity 1.2 is that cognitive psychologists try to devise studies that isolate the particular

BOX 1.3 RESEARCH STUDY Category-specific impairments III: cognitive modelling

Greer *et al.* (2001) developed a computational model based on the assumption that living things and non-living things are not represented in qualitatively distinct ways, but that differences between them arise because living things have many shared properties that are strongly correlated (all mammals breathe, have eyes, etc.), whereas the properties of non-living things tend to be more distinct. Greer *et al.* developed a form of computational model called a connectionist network, which encoded these differences between living and non-living things. The model contained three kinds of units organized in three layers, as shown in Figure 1.3.

However, information about the categories was distributed over the network's units in such a way that it was not possible to associate individual units with either living or non-living things. Greer *et al.* then artificially lesioned or damaged their network by removing 10 per cent of the network's connections at a time. They found that the shared properties of living things were more impervious to damage than those of non-living things, as shown in Figure 1.4. Of course, debates concerning the appropriate interpretation of category-specific impairments continue – see, for example, Martin (2007) and Moss *et al.* (2009).

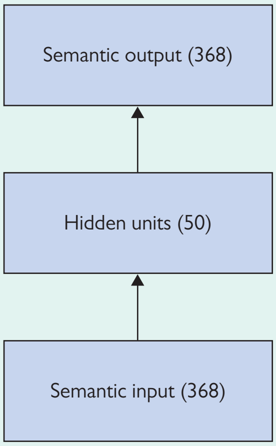


FIGURE 1.3 Architecture of Greer *et al.*'s connectionist network. The semantic input layer represents properties of categories. The network was trained until it could reproduce in the output layer the same pattern presented to its input layer. Arrows imply that every unit in a layer is connected to every unit in the subsequent layer. Numbers indicate the number of units in each layer. *Source: Tyler and Moss, 2001, Figure 1 p.248*

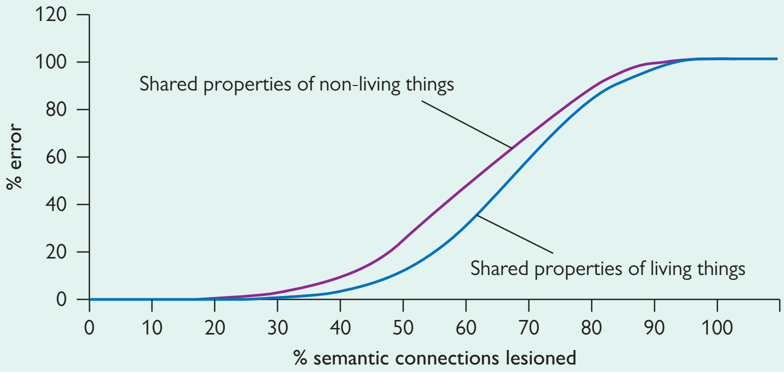


FIGURE 1.4 The results of 'lesioning' the model to simulate brain damage. As predicted by Greer *et al.*, the shared properties of living things were better preserved than the shared properties of non-living things, owing to the greater correlations between them. *Source: Tyler and Moss, 2001, Figure 11, p.249*

✱ ACTIVITY 1.2

At this moment your behaviour involves getting information from this book. Your eyes may be scanning across the page and detecting patterns of colour and light and shade, or, if you are listening to this book on audio CD or it is being machine-read from an electronic copy, your ears will be detecting sound waves of varying intensity and pitch. Your behaviour can also be seen in a wider context: it is just one aspect of what is involved in studying psychology. Take a few minutes to jot down your explanation for your behaviour – if someone were to ask why you are behaving in this way, what would your answers be? Try to think of as many different ways of answering the question as you can. Also list any processes that you think might be going on in your mind – how would you describe them?

COMMENT

The first thing to note is that your behaviour can be explained in many different ways. For example, you might have noted that your reading is bound up with a *feeling of elation* – perhaps you love studying cognitive psychology – or a *feeling of anxiety* – perhaps you are uncertain of obtaining a good course grade. Your explanation adverts to *emotions*. Perhaps you jotted down as an answer that you *reasoned* that you ought to read this book since you want to do well on your course. Perhaps doing well on your course is part of a strategy to reach a goal, or *solve a problem* such as how to improve your qualifications. You might also have suggested that you *decided* to read this book – perhaps faced with different ways of spending your time, you *judged* that this would be the most beneficial (we'll try not to let you down!). You might have thought that there are processes going on in your mind to do with *reasoning*, *problem solving*, and *decision making*.

It might be that you are reading this chapter for a second time because you want to make sure you *remember* it. So, your explanation adverts to *memory* and the processes that are responsible for things being remembered (and forgotten).

How else might you have explained your behaviour? You might have suggested that you were trying to *understand* the chapter; that you behaved the way you did because you were involved in understanding words, phrases, and sentences. You may have indicated that there must be processes for understanding *language*. Perhaps there were other explanations you offered. Maybe you explained your reading of the book by saying 'That is what books are for' – because you *categorized* it as a book. Maybe you suggested you were scanning your eyes across the page in order to *perceive* and *recognize* words. And, just maybe, you suggested that your behaviour was happening because you were paying *attention* and not being distracted by a telephone or a doorbell.

The words in emphasis in the previous paragraphs all provide important means for explaining behaviour that are used by cognitive psychologists, and are all major topics of this book.

cognitive processes under investigation. For example, a researcher interested in language processing will try to devise their studies so that they measure language processes only and are not unwittingly influenced by other processes, such as emotion or reasoning. Consider also how the studies referred to in Boxes 1.1 to 1.3 try to focus exclusively on the issue of category specificity. Indeed, it is a general strategy within cognitive psychology to try to isolate particular cognitive processes for further investigation. Table 1.1 lists some prevalent assumptions that this strategy gives rise to.

TABLE 1.1 Assumptions commonly made in the cognitive approach

1	Cognitive capacities can be partitioned so that individual capacities can be studied in isolation (e.g. so that language can be studied in isolation from memory)
2	Cognitive psychology tends to focus on the individual and their natural environment (relatively de-emphasizing the roles of culture and society)
3	Cognitive capacities are relative autonomous when compared with non-cognitive capacities (e.g. affect, motivation, etc.)
4	It is useful (and meaningful) to distinguish 'normal' from 'abnormal' cognition
5	Adults are sufficiently alike that we can talk of a 'typical' cognizer, and generalize across cognizers, ignoring individual differences
6	Answers to basic, empirical questions can be given in terms of information processing
7	Answers to basic, empirical questions should be justified on empirical grounds
8	Answers to the basic, empirical questions must be constrained by the findings of neuroscience (as and when these are relevant)

Source: adapted from Von Eckardt, 1993, pp.54–5

SUMMARY OF SECTION 2

- Cognitive psychology can be characterized as the scientific study of the mind.
- Cognitive psychology can be characterized in terms of its methods:
  - experimental studies of normal cognition
  - neuropsychological studies that relate normal to abnormal cognition
  - neuroimaging studies that reveal the location and/or the time course of brain activity
  - computational models that can be tested and compared with experimental data.
- Cognitive psychology can be characterized in terms of its subject matter (see the table of contents for this book).
- Everyday behaviour involves multiple cognitive processes:
  - cognitive studies tend to isolate one process or set of processes for study.

3 A BRIEF HISTORY OF COGNITIVE PSYCHOLOGY

Cognitive psychology did not begin at any one defining moment and there are many antecedents to its evolution as a branch of enquiry. In this section we will briefly sketch some of those antecedents and try to indicate how and why they resulted in the development of what today we call cognitive psychology. However, all written history is necessarily selective and simplified, and a historical account as brief as the one we are about to give must be especially so. We start with introspectionism.

3.1 Introspectionism

Modern experimental psychology has its roots in the work conducted in Europe in the mid-nineteenth century by such people as Donders, Fechner, Helmholtz, and Mach. When Wundt established the first dedicated psychology laboratory in Liepzig in 1879, he sought to build upon the efforts of these pioneers. He took *consciousness* to be the proper subject matter of psychology. According to Wundt, physical scientists study the objects of the physical world either directly or, more often, through observation of the readings on instruments. In either case, observation is mediated by conscious experience, but for physical scientists things in the world are the object of study, not the conscious experience by means of which we know them. Psychology would be different in that it would take as its subject matter conscious experience itself.

Wundt adopted **introspection** as a research method, believing that properly trained psychologists should be able to make observations of their own experience in a manner similar to the way that properly trained physicists make selective observations of the world. Wundt fully understood the need to design experiments with adequate controls and produce replicable results. He also made use of objective measures of performance, such as reaction time (RT). The focus of his interest, however, was the conscious experience that preceded the response. For example, if one condition in an experiment yielded longer RTs than another, he wanted to know how the two preceding conscious experiences differed. Wundt was not concerned with the unconscious processes involved in responding to a simple stimulus – the rapid information-processing operations that, as you will find in the following chapters, form much of the subject matter of modern cognitive psychology. He considered these to lie in the realm of physiology rather than psychology.

In opposition to Wundt's Liepzig school was the Würzburg school of introspection. Its leader, Külpe, was a former student of Wundt's, who with his colleagues and students developed an alternative view of conscious experience and what could be revealed by introspection. We can characterize the main difference between the two schools in terms of a distinction that will be more fully introduced in Chapter 3 in relation to the topic of perception, although the protagonists would not have used these exact terms themselves. Put simply, the Liepzig school held that the contents of consciousness are constructed 'bottom-up' from simple sensations combined in accordance with the strength of associations between them (something like the connectionism you can read about in Chapters 4, 19, and 20, and elsewhere). The Würzburg school, on the other hand, held that the contents of consciousness are determined in a much more 'top-down' fashion by the nature of the task that one is engaged upon. Külpe and his colleagues sometimes studied simple tasks, but tended to favour more complex ones in which mental acts such as attending, recognizing, discriminating, and willing played a larger role.

Introspectionism went into a terminal decline during the first two decades of the twentieth century. The details of the many unresolved disagreements between the two schools of introspectionism need not detain us here, but it is worth noting two things. First, the introspectionists developed elaborate classifications of conscious experience, a topic that has quite recently begun to attract the attention of psychologists once again (see Chapter 18). Second, although psychologists began to lose interest in consciousness during those two decades, the exploration of consciousness still remained central to developments in the visual and literary arts (e.g. cubism and expressionism in painting, and James Joyce, Virginia Woolf, and Gertrude Stein in literature).

3.2 Gestalt psychology

The perceived failures of introspectionism provoked a number of intellectual reactions. In Europe, the Gestalt psychologists built upon the work of the Würzburg school and argued that the contents of consciousness cannot be analysed into simple component sensations. According to Wundt, the perception of movement results from a sequence of sensations corresponding to an object occupying successive locations over time. However, Wertheimer argued in 1912 that 'pure movement' can be perceived directly; it does not have to be 'inferred' from changes in the location of an object. A good example is when we see the wind gust through grass. Blades of grass bend in succession but no blade changes location. What we perceive is pure motion (of the invisible wind) without a moving object. (Modern studies show that motion perception can, in fact, arise either on the basis of the changing location of an object or from successive changes across space without a moving object.) Gestalt psychologists also emphasized the importance of the perception of stimulus patterning to our conscious experience. A tune played in one key on one sort of instrument remains the same tune when played in another key or on a different instrument. Since the notes, or the sounds making up the notes, have changed, there must be more to the tune than can be found by an analysis into simple auditory sensations. The tune is in the perceived relationships between the notes, their patterning.

Meanwhile, in the USA, William James opposed introspectionism with his '**functionalist psychology**'. Sounding remarkably like an exponent of what is now called evolutionary psychology, James stated that, 'Our various ways of feeling and thinking have grown to be what they are because of their utility in shaping our reactions to the outer world'. These functions of the mind were, in James's view, the proper subject matter for psychology. Perceiving and thinking, grief and religious experience, as psychological functions, were themselves to be the focus of interest, rather than the evanescent contents of consciousness on which the introspectionists had fixated. However, James's ideas were soon to be largely swept aside by another more powerful current in US thought: behaviourism.

3.3 Behaviourism

The founders of behaviourism were driven by various motives, not all shared in common. Watson, the principal standard-bearer for the new kind of psychology, was especially keen to move psychological research out of the laboratory and into 'the real world'. He was less interested in fine distinctions of conscious experience than in how people act in everyday life, and how they can be influenced. He wanted to see psychological knowledge applied to education, clinical problems, and advertising, and he initiated work in all these areas. Not all behaviourists were as zealous as Watson when it came to applying psychology, but one belief they did have in common was that psychology should be scientific and objective; by this they meant that its subject matter should be as publicly observable as possible. Consciousness is (at best) only privately observable. What can be publicly observed are stimuli, behaviour, and reinforcers. So psychologists such as Thorndike, Watson, and later Skinner, Eysenck, and others argued that psychology should be scientific in its approach, and should seek to explain behaviour through reference mainly to such observables. We say 'mainly' because it is impossible to convey briefly all the complexities and nuances of the varied schools of behaviourism. For example, Eysenck's theory included individual differences in a biologically rooted propensity to seek or avoid stimulation. Nevertheless, the behaviourist emphasis on public observation was intended to place psychology on an objective footing, akin to the natural sciences like physics and chemistry, and it reflected a wider philosophical consensus as to the proper nature of scientific enquiry.

3.3.1 Science and the unobservable

In all human efforts to comprehend the world there is a tension between on the one hand observable events, and on the other hand the often encountered need when explaining them to postulate unobservable theoretical entities and forces, whether gods or atoms. This tension is central to science. A key idea in the development of science has been that knowledge should be empirical, based on experience and not on received wisdom or purely rational calculation. Observation is one of the touchstones of science, but scientific theories also refer to unobservables. The explanation that physics offers for an apple falling to Earth invokes the notion of a gravitational force, something that is not directly observable. Similarly, in explaining why a compass needle points to magnetic north, physicists talk of magnetic fields and lines of magnetic force. But these things too are unobservable. If you have ever placed iron filings near a magnet you will have seen that they move to orient themselves along the lines of the magnetic field. But, strictly, we don't observe the magnetic field or the lines of magnetic force, but rather their influence upon the iron filings. All natural sciences employ unobservable, theoretical constructs that are invoked in order to explain observations. For example, chemistry appeals to notions such as the energy levels of electrons in order to explain why compounds react. These levels are unobservable too, of course. So, the fact that a discipline is committed to explaining observed behaviour by reference to hypothesized, unobservable constructs does not in itself render the discipline unscientific.

However, to find scientific acceptance, unobservable constructs have to be seen to do useful theoretical work. When Newton proposed the notion of a gravitational force, certain critics immediately accused him of introducing a mystical notion into 'the new science'. Newton's ideas gained acceptance only because they met other scientific criteria – such as elegance, simplicity, and rigour – and because the concept of gravitation, despite its somewhat mysterious nature, had a wide range of applications. Gravitation explained not just the fall of objects to the ground but also the rhythm of the tides and the movements of the planets. It could also be precisely formulated mathematically as an inverse square law: the attraction between any two bodies is proportional to the square of the distance between them. In other words, the willingness of the scientific community to countenance a hypothetical unobservable depends on how useful it is judged to be on a range of criteria.

Science has had to live with the necessity for unobservables. But acceptance through necessity is not the same as liking, and science always receives a boost when a technical breakthrough brings a previously unobserved entity into the realm of observation for the first time. For example, Mendel postulated 'units of heredity' on the basis of his plant-breeding observations, but these ideas were felt to be on a firmer footing once new technology made it possible to see chromosomes and genes. Thus, scientists are forced somewhat grudgingly to accept the need for postulating unobservables. And because science – like all human institutions – is subject to swings of fashion, the willingness to countenance unobservable theoretical entities fluctuates over time. For reasons that we are unable to describe here, but that were rooted in the growing crisis of classical physics that would culminate in the birth of quantum theory and relativity theory, the late nineteenth and early twentieth century was a period during which scientists were particularly intolerant of unobservables. The importance of observation became enshrined in the assumption known as operationism. This is the idea that theoretical concepts are only meaningful to the extent that they can be exhaustively analysed in terms of things that can be observed.

3.3.2 Back to behaviourism

The bias against unobservables affected all the traditional sciences and also the newer, aspirant scientific disciplines such as physiology and psychology. The introspectionists, with their 'observations' of consciousness, had responded to it, but the intellectual climate seems to have been especially suited to propagating an emphasis on what could be publicly observed. With the decline of introspectionism, behaviourism was taken up enthusiastically, first in the USA and then more widely.

While behaviourists could, perhaps, concede the *existence* of consciousness while arguing that it was not appropriate for scientific study, at least some of them felt that operationism committed them to the stronger claim that talk of consciousness was not even meaningful. Of course, behaviourism has never been a single view, and since the time of Watson and Thorndike behaviourists of various hues have modified their positions. Skinner, for example, conceded that internal mental events, including conscious experiences, might *exist* (indeed they were construed as forms of covert behaviour). But despite this rejection of operationism, even Skinner still thought that talk of internal events should be avoided within a scientific psychology.

You might think that avoiding talk of internal events might make it impossible to explain many, or even most, psychological phenomena. However, behaviourists were concerned to show how even complex phenomena might be understood in terms of principles of learning, with behaviour seen as being made up of learned responses to particular stimuli. One view of language production, for example, was that the utterance of a word could be seen as a learned response. The utterance of a whole sentence could be seen as involving a chain of stimulus–response pairs, in which each response (the utterance of a word) also serves as the stimulus that leads to the production of the next response (the next word).

Despite the possibility of giving behaviourist explanations of complex activities such as the utterance of a sentence, behaviourists tended not to offer accounts of what we now refer to as higher mental processes – processes such as producing and understanding language, planning, problem solving, remembering, paying attention, consciousness, and so on. As the years passed, however, some psychologists came to see this as a major failing.

3.4 The return of the cognitive

In 1948, at a meeting known as the Hixon symposium, Karl Lashley gave a talk entitled 'The problem of serial order in behaviour' (Lashley, 1951). In this, he gave prominence to the problems posed for behaviourist accounts by complex actions in which behaviour segments are somehow linked together in a sequence, and where two segments depend upon one another even though they may be separated by many intervening segments. Language, as you might have guessed, provides a prime example. In fact, the last sentence illustrates the point nicely: when I came to write the word 'provides' in the previous sentence I chose to end it with the letter 's'. I did so, of course, because this verb has to agree grammatically with the singular noun 'language', the subject of the sentence. In my actual sentence, these two words were separated by a clause, and so my action at the time of writing the word 'provides' depended upon a much earlier behaviour segment – my writing of the word 'language'. Lashley argued that since the production of some words in a sequence could be shown to depend upon words produced much earlier, the simple view that each word is the stimulus that produces the subsequent word as a response could not properly explain language production.

He also argued that many behaviour sequences are executed simply too rapidly for feedback from one segment to serve as the trigger for the next. He cited examples such as the speed with which pianists and typists sometimes move their fingers, or with which tennis players adjust their whole posture in response to an incoming fast service. Lashley's alternative to the chaining of behaviour segments was to suppose that complex sequences are planned and organized in advance of being initiated. The speech errors discussed in Chapter 7 of this book provide especially compelling examples of the kind of planning and organization that underlie skilled behaviour.

Lashley's view that behaviourism could not properly explain how people produce (or comprehend) language was later reinforced by a review of Skinner's book *Verbal Behavior* (1957) by the linguist Noam Chomsky (1959). Chomsky argued, contra behaviourism, that language could not be thought of as a set of learned responses to a set of stimulus events. His argument had a number of different aspects. For example, he argued that children seem to acquire their first language too effortlessly – if you have tried to learn a second language you can perhaps testify to the difference between learning a first and learning a second language. While the latter seems to require intensive and effortful study, the former is something that pretty much everyone does without the need for formal schooling. He also argued that if the behaviourists were right, then exposing children to impoverished or ungrammatical language should hinder their learning of the correct stimulus–response relationships. Yet studies show that much of the speech to which young children are exposed is indeed ungrammatical and otherwise impoverished, and this in no way prevents them from learning the grammar of their native tongue. Similarly, Chomsky argued that general intelligence ought to influence the learning of stimulus–response relationships. Again, however, intelligence does not seem to influence whether or not children learn the underlying grammatical rules of their language. Chomsky presented many other arguments to the same effect, and though many of these have been thought to be contentious, his position was extremely influential in setting up an alternative, cognitive conception of language. Most significantly, Chomsky proposed that language is rule-based and that, far from children learning language by learning how to respond to particular stimuli, their acquisition of language involves acquiring its rule-base. On this view, my being able to write grammatical sentences involves deploying my (generally implicit, or unconscious) knowledge of the rules of language. In referring to such implicit knowledge, Chomsky proposed that an understanding of how people produce, comprehend, or acquire language will necessarily involve reference to something that cannot be directly observed – their knowledge of the underlying rules, or organization, of the language.

Although this emphasis on the role of planning, organization, and rules in the generation of behaviour was to be hugely influential from the 1950s onwards, these ideas were certainly not new to psychology. As mentioned previously, the Gestalt psychologists had drawn attention earlier in the century to the importance of patterning, or organization, for perception, and the same point was also made in relation to action. Someone who has learned to sing or hum a tune can very probably manage to whistle it thereafter. Yet singing, humming, and whistling call for very different sequences of muscle movements. This indicates that learning a tune must involve learning a set of abstract relationships between notes, which can be instantiated as any of a variety of muscular productions. A similar idea, that what is learned must often be more abstract than straightforward stimulus–response connections, was also expressed by the school of 'cognitive behaviourists' associated with Tolman (1932). Rats that had learned, for example, to turn left repeatedly in a maze to find food were shown to swim left when the maze was flooded. Since the muscle movements of running and swimming are completely different from one another, the rats must clearly have learned something more abstract than a particular chain of muscular responses.

Even before the writings of the Gestalt psychologists or the work of Tolman, psychologists studying the acquisition of skills had realized the importance of planning and organization for the production of skilled behaviour, such as in Morse telegraphy or typing (Bryan and Harter, 1899). At the time of the Hixon symposium, therefore, there were already existing traditions within psychology upon which the renewed interest in the planning and structure of behaviour could draw. And of course, the intellectual climate of the mid-twentieth century was changing rapidly in many other ways too. New technologies were influencing the ability of scientists to conceptualize the workings of complex systems. One of the most crucial issues related to the type of causal explanation that is appropriate to explain the behaviour of such a system. Purposive, or teleological, explanations had been taboo in Western science since the time of thinkers such as Galileo and Newton. Where, for example, an ancient Greek philosopher might have said that a stone falls to Earth 'in order to' reach its natural resting place at the centre of the Earth (which was also the centre of the Greek universe), Newton said that the stone falls because it is acted upon by the force of gravity. The strategy of explaining phenomena in terms of causes that precede and 'push' their effects, rather than in terms of goals or final states towards which events are 'pulled', had proved highly successful in the physical sciences. The move from goal-directed, purposive explanations to mechanical cause-effect explanations was usually considered to be a move from pre-scientific, animistic thinking to proper scientific thinking. Behaviourism was, and still is, an attempt to bring psychology into step with this way of analysing phenomena. A strict emphasis on an organism's history of conditioning allows an explanation of behaviour in terms of prior causes rather than future goals. However, the development of progressively more complex artificial devices started to call into question the universal applicability of explanations in terms only of prior causes. It became increasingly clear that, while the functioning of the mechanical parts of any such system *can* be explained in cause-effect terms, such explanations will never capture the function (or purpose) of the whole system.

Central to the new kind of apparently purposive machines (known as servomechanisms) was a reliance on feedback loops. **Feedback** is information about the match or mismatch between a desired goal-state and an existing state of affairs. The classic example is the domestic central heating system, in which the thermostat setting selected by the householder is the goal-state and the temperature measured by an air thermometer is the existing state. The two are compared mechanically. If the existing temperature is less than the desired temperature, this negative feedback is transmitted to the boiler controls causing the boiler to be switched on.

The boiler continues to fire until information has been fed back to the boiler controls that the discrepancy between the actual and desired temperatures has been eliminated. The system as a whole exhibits a simple but dynamic behaviour, with the boiler turning on and off in a manner that maintains room temperature at or about the desired level. Importantly, the function of maintaining a steady temperature cannot be localized to any one component of the heating system, such as the thermostat, the thermometer, the boiler, or its controls, but is a property of the system – as a whole.

Far more complicated servomechanisms with more complex feedback controls were also being developed. Anti-aircraft gunnery may not seem very pertinent to an understanding of animal and human behaviour, but it was partly as a result of working on gunnery problems in the Second World War that the mathematician Norbert Weiner developed the notion of 'cybernetics', the science of self-governing, or goal-directed, systems. Accurate anti-aircraft gunnery requires that a projectile is fired and timed to explode not at the present location of the target aircraft, but at its future location. This means not only predicting the future position of the plane but also rotating the gun so it faces in the appropriate direction and with the correct elevation. Clearly, humans successfully extrapolate flight paths and aim at future positions when, for example, shooting game birds. However, for planes flying at ever greater heights and speeds, calculation of the necessary trajectory of the projectile exceeds human capabilities and must be computed automatically. Moreover, using motors to move a gun weighing many tons is a very different matter from moving a shotgun, or indeed a bow and arrow, held in your arms. Although we are mostly unconscious of it, normal bodily movement is based upon continuous muscle, tendon, and visual feedback about how the movement is proceeding. Unless similar feedback is designed into the gun control system, the swinging anti-aircraft gun may easily undershoot or overshoot the intended position, particularly as, depending on the air temperature, the grease packed round the mechanism will be more or less 'stiff'. Apply too little power and the gun will undershoot the intended position, a second push will be required, and the gun will 'stutter' towards its position. Apply too much force and the gun will overshoot and will have to be pulled back, in what can turn into a series of increasingly wild oscillations. Engineers discovered that the smoothest performance was achieved by using feedback loops to dynamically control the turning force applied to the gun.

Weiner, and other cyberneticists such as Ashby, recognized the importance of feedback and self-correction in the functioning of these new and complex technological devices, and they also saw analogies with complex natural systems. Weiner drew parallels between the effects of certain neurological conditions and damage to the feedback control of behaviour. For example, the tremors observed in Parkinsonian patients were likened to the oscillations of an anti-aircraft gun when its movement is insufficiently 'damped' by feedback control.

An important intellectual leap for cognitive psychology came with the realization that just the same kind of analysis can be applied at any level of behavioural control. In other words, it is not just automatic homeostatic functions or unconsciously executed movements that can be analysed in terms of feedback loops, but any function/behaviour from the wholly non-conscious to the fully conscious and intended. Miller *et al.* (1960) developed the notion of feedback control into the hypothesis that behaviour (of animals, humans, or machines) can be analysed into what they called **TOTE units**. TOTE stands for Test-Operate-Test-Exit. A test is a comparison between a current state and a goal-state. If a discrepancy is registered, some relevant operation intended to reduce the discrepancy will be performed (e.g. switch on the boiler). A second test, or comparison, is then conducted. If a discrepancy remains, the operation can be repeated, followed by another test. If the discrepancy has been eliminated, the system exits the TOTE unit.

Miller *et al.* conceived of the TOTE unit as an advance on the conditioned reflex notion of Pavlov and the conditioned response notion of Watson and Skinner, both of which can be conceptualized as TOTEs. The aim was to develop a unit of analysis of behaviour that could apply to everything from a dog's conditioned salivatory response to deliberate, planned action. The TOTE provides a basic pattern in which plans are cast; the test phase specifies what knowledge is necessary for a comparison to be made, and the operation phase specifies what the organism does about the outcome of the comparison. Although this scheme makes it possible to talk about purposive behaviour and about unobservable goals and comparison operations, there is continuity from behaviourism. Cognitive psychology generally attempts to retain the scientific rigour of behaviourism while at the same time escaping from the behaviouristic restrictions in relation to unobservables.

An important property of TOTEs is that they can be nested within hierarchies. The operation segment of any TOTE can itself be composed of one or more TOTE units. For example, the TOTE for starting the car might be nested within the operation of a larger TOTE for driving to the shops, which might itself be nested within a still larger unit having the goal of buying a present. This nesting of feedback loop units provides a way to conceptualize how behaviour can be complexly structured. In this scheme, moment-to-moment control of behaviour passes in sequence between a series of TOTE goal-states, with the TOTE units themselves nested in hierarchies. Miller *et al.* explicitly likened this 'flow of control' of behaviour to the way in which control in a computer program switches in orderly fashion from command line to command line as the execution of any particular sub-routine is completed. (Note: what 'flows' around a TOTE can be energy, information, or, at the highest level of conceptual abstraction, control.)

3.4.1 Computers and the mind

Another development in the mid-twentieth century with a huge import for the development of cognitive psychology was the opening up of a new field concerned with the possibility of designing and then building computers. Building on earlier work that developed a formal or mathematical approach to logical reasoning, Claude Shannon in 1938 showed how core aspects of reasoning could be implemented in simple electrical circuits. In the 1940s, McCulloch and Pitts showed how it was possible to model the behaviour of simple (and idealized) neurons in terms of logic. Taken together, these developments suggested something that at the time seemed extraordinary – that the brain's activity could, at least in principle, be implemented by simple electrical circuits.

In parallel with these developments, the 1930s and 1940s saw pioneering theoretical developments in computation and information processing. Turing, in 1936, developed an abstract specification for a machine (a Turing machine) that could compute any function that in principle could be computed. In the 1940s, Shannon and Weaver used the tools of mathematics to propose a formal account of information and of how it could be transmitted.

Technological progress was also rapid. In 1941, Konrad Zuse of Berlin developed the world's first programmable, general-purpose computer. In 1943, Colossus, a special-purpose computer designed to break wartime codes, became operational at Bletchley Park in Buckinghamshire. In 1946, John von Neumann articulated a set of architectural proposals for designing programmable, general-purpose computers. These were adopted almost universally and computers have since also been known as von Neumann machines. In 1948, the Manchester University Mark I programmable, general-purpose computer became operational, and in 1951, Ferranti Ltd began producing, selling, and installing versions of the Manchester Mark I – the world's first commercially available, programmable, general-purpose computer.

These developments, fascinating though they were in their own right, also seemed to carry important implications for our understanding and study of the mind. They appeared to show, for instance, that reasoning, a central feature of the human mind, could be implemented in a digital computer. If that were the case, then not only could the computer be used as a tool to aid our understanding of the mind, but the question would also arise as to whether minds and computers are essentially alike. Indeed, in 1950 Turing proposed a test – the Turing test – by which he thought we should judge whether two entities have the same intelligence. Turing believed that, should the situation ever arise whereby we could not distinguish the intelligence of a human from the 'intelligence' of a computer, then we ought to concede that both were *equally* intelligent. Moreover, since we are in agreement that humans are capable of thought, we ought to concede that computers are also capable of thought! Box 1.4 outlines the Turing test and considers what it might take for it to be passed.

BOX 1.4 RESEARCH STUDY The Turing test: can computers think?

Turing proposed that we could determine whether a computer can think by judging whether it succeeds in what he called the imitation game. In the game there are three participants, two humans (*A* and *B*) and a computer (*C*). The arrangement of the participants and the communication flow between them is schematically indicated in Figure 1.5.

The participants are positioned in separate rooms, so each one is unable to see, hear, or touch the others. However, one of the human participants (*A*) is connected via a VDU terminal connection to the other human participant (*B*) and also to the computer (*C*). *A* can communicate electronically with both *B* and *C*. The goal for *A* is to ascertain which of *B* and *C* is the computer and which is the human. The goal of *B*, the other human, is to assist *A* in making the correct identification (perhaps by trying to appear as human as possible). *C*'s goal, by contrast, is to lead *A* into making the wrong identification (by imitating human behaviour). *C* wins the game if *A* cannot reliably identify *C* as the computer. Turing's claim was that if a computer could simulate human behaviour so successfully that another human could not tell that it was a computer, then the computer could legitimately be said to think.

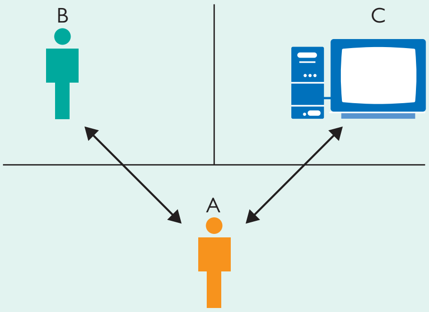


FIGURE 1.5 The arrangement of the participants in Turing's imitation game.

Turing's position remains controversial, of course, though it certainly captured the imagination of the time. In 1956, at the Dartmouth Conference (held in Dartmouth, New Hampshire), John McCarthy coined the phrase 'artificial intelligence' (or AI). He founded AI labs at MIT in 1957 and then at Stanford in 1963, and so began a new academic discipline, predicated on the possibility that humans are not the only ones capable of exhibiting human-like intelligence.

You have now been introduced to a variety of the influences that go to make up cognitive psychology. Cognitive psychology inherits some of the behaviourist concerns with scientific method. Throughout this book you will see that almost constant reference is made to systematic observations of human behaviour (and sometimes animal behaviour too). Almost every chapter will present the results of empirical investigations, and these are fundamental in guiding our understanding. But cognitive psychology rejects the exclusive focus on what is observable. As Chomsky implied, understanding the mind requires us to consider what lies behind behaviour – to ask what rules or processes govern the behaviour we observe. Each chapter will also consider the extent to which we understand how the mind processes information, and how that information is represented. Cognitive psychology also has a major commitment to the use of computers as devices for aiding our understanding of the mind. First, computers are used as research equipment to control experiments, to present stimuli, to record responses, and to tabulate and analyse data. Second, computers are also used as a research tool – if we can implement reasoning in a computer, for example, we may gain insight into how reasoning might be implemented in the brain. So, most of the chapters in this book will also discuss ways in which researchers have used computer models to help us understand how the mind processes and represents information when people perform certain behaviours. Third, and more controversially, computers are also considered to be candidate 'thinkers' in their own right. Understanding more about the nature of computation itself may shed light on the nature of thinking and on the nature of the mind.

SUMMARY OF SECTION 3

- Cognitive psychology inherits some of the behaviourist concerns with scientific method. Almost every chapter in this book presents the results of empirical investigations, investigations that are fundamental in guiding our understanding.
- Cognitive psychology rejects an exclusive focus on what is observable. Almost every chapter considers the extent to which we understand how the mind processes information and how that information is represented.
- Cognitive psychology is committed to using computers as a tool for aiding our understanding of the mind.
- Introspectionist and Gestaltist interest in conscious experience was replaced by the behaviourist focus on what is publicly observable.
- There is always a tension in science between the emphasis on observation and the need to postulate unobservable theoretical entities.
- Behaviourists did not necessarily deny the importance of higher mental functions but rarely offered accounts of them.
- Cognitive psychology has many roots; it has been heavily influenced by technological developments and the way they help us to understand complex behaviours.

4 SCIENCE, MODELS, AND THE MIND

If cognitive psychology is concerned with the processes and representations of the mind, and these cannot be directly observed, how can cognitive psychologists bridge the gap? How do we speculate about the nature of something we cannot observe, while remaining scientific? There are broadly three kinds of answer.

First, as we have already discussed, scientific theories commonly invoke unobservable theoretical entities to account for observational data (e.g. force fields, electron energy levels, genes, or cognitive operations).

The second answer builds on the first. When a theory hypothesizes an unobservable, theoretical construct, a model needs to be specified of the relationship between the construct and the behaviour to be explained. It would have been insufficient for Newton to have tried to explain why things fall to Earth by simply invoking the notion of gravitation. He went further and derived equations to model the effects of gravity, which can be used to generate predictions about how gravity ought to work for things whose motion has not yet been systematically observed. So physicists could then perform studies in order to confirm the predictions (that is, until Einstein's theories of relativity, but that is another story).

Cognitive psychology proceeds in a similar way. Consider again the example of language. Cognitive psychologists have made numerous detailed observations of the production (and comprehension) of language (you can find discussions of these in Chapters 6 and 7). Explaining these observations, however, seems to require positing things internal to the mind that are involved in producing the observed behaviour. These are the unobservable, theoretical constructs of mental processes and structures. Positing these, of course, is just the starting point. The challenge for cognitive psychologists has been to say more. They have to develop models of these mental structures and processes, show how they give rise to the observed behaviour, and, importantly, show how successfully they predict behaviour that has not yet been systematically studied in experiments.

Developing a model is not easy; Newton apparently needed the inspiration provided by an apple falling to Earth (or so the story goes). And much of the challenge facing cognitive psychologists is to harness their creativity and imagination in order to suggest plausible models. Throughout your reading of this book, you might wish to consider how you would have responded to some of the problems described. You might want to consider what would constrain your choice of model, what kinds of model you would have developed, and how you would have set about doing this. Without doubt, these are difficult questions – so don't lose too much sleep over them! – but they at least serve to show how creative cognitive psychology is. The matter of devising studies in order to evaluate a model is also creative. By working out the predictions a model might make, psychologists can evaluate it by devising studies to test its predictions and by then making the relevant behavioural observations.

Creating models and designing studies to test them is not easy, but cognitive psychologists can use computers to help. The previous section suggested two ways in which computers are important to cognitive psychology other than as experimental equipment – computers might be capable of thought, and they can also serve as tools for implementing models such as a model of language processes. Now, perhaps, you can see how they might contribute to the scientific objectives of cognitive psychology – researchers can use computers in order to create models. Just as computer programmers can build programs to do things such as word processing or financial accounts, so researchers in cognitive psychology can program computers to behave according to a particular model of the mind. Using computers to program particular models can be helpful on a number of counts:

1. Models can rapidly become very complicated – too complicated to be expressed verbally, or for one person to hold all the relevant details in mind. This problem affects others too – meteorologists increasingly use computer models of weather systems and economists use computer models of the economy. The phenomena involved are so complicated that, without computers, they would be almost impossible to model.
2. It is not always easy to work out the predictions of a model. Programming a model can allow researchers to simulate the effects of different conditions and so find out how the model behaves, and whether this behaviour accurately predicts how humans will behave.
3. Perhaps most important of all, by programming a model into a computer researchers can determine whether the model is internally consistent (whether there are statements in the model that contradict one another) and whether the model is already clearly and precisely stated. If it is, the computer program will run; otherwise, it will crash.

So cognitive psychology can posit the existence of unobservable (cognitive) processes and structures and still be scientific. Not only is this true of other disciplines like physics and chemistry but, like those disciplines, the gap between observable behaviour and unobservable processes and structures can be bridged via the creation and evaluation of models.

There is, however, a new possibility for linking cognitive processes with a focus on observation, and this leads to the third answer to the question with which this section began. The advent of new techniques for imaging the brain suggests that, just possibly, mental processes and structures may not be entirely unobservable (as the behaviourists once believed).

Functional MRI studies (and other kinds of imaging) allow us to see which parts of the brain become especially active when people are engaged in a certain task (relative to when they are engaged in some control task or tasks). There is considerable debate in the cognitive community as to the usefulness of imaging techniques for helping researchers to develop theories of cognition. Activity 1.3 will help you get a sense of the issues involved.

✱ ACTIVITY 1.3

Consider again the brain images in Figures 1.1 and 1.2. First, think about what you could infer from the images alone. What does the indication of activity in particular brain regions tell you? Second, think about the processes going on inside participants' minds. What additional information would you need in order to be able to say what the brain activity represents? Suppose you were given very detailed anatomical descriptions of the active regions – what would that enable you to conclude?

COMMENT

It is one thing to say that there is activity in particular regions of the brain, but quite another to say exactly what cognitive processes and structures are involved. An image of brain activity, on its own, does not help very much. Seemingly, what is crucially needed is further evidence as to what information each brain region processes. That is, we need to know the function of the active regions. One way of trying to identify the function of different brain regions is to compare brain images for different kinds of task – regions that are active for all tasks may be implicated in information processing that is common to those tasks. This assumes we have good models for the information-processing characteristics of different tasks. If so, and also using anatomical and neuropsychological evidence, researchers can then tentatively begin to identify particular regions with particular functions. This in turn can help researchers to interpret and design further brain-imaging studies.

One criticism of imaging studies is that, at best, they help researchers to localize a particular function – that is, researchers can identify the function with a particular region of the brain – but that they do not improve our theories of cognition. However, this is a bit like saying that being able to see chromosomes and genes down a microscope does not improve the theory of genetic inheritance. In one sense that is true, but making visible entities that were previously only theoretical does increase overall confidence in the theory. Similarly, suppose a cognitive theory says that reading some words involves using a visual processing route and reading other words involves using an auditory processing route. Finding that the first task induced activity in areas known to be engaged by other visual tasks and that the second task induced activity in areas known to be engaged by auditory tasks would increase our confidence in the theory.

Without prejudging the ongoing debate in this area, it is likely that imaging techniques will contribute to cognitive theory in various ways. Sometimes the contribution will be at the level of theoretical deduction; sometimes it may be at a less palpable level, as when it adds to the confidence in a theory. When genes were first made visible, genetic engineering was a very distant prospect, but it is hard to imagine the latter without the former. The advances in cognitive sciences to which neuroimaging will contribute are equally hard to predict but we shall be surprised if they do not prove to be many and varied.

SUMMARY OF SECTION 4

- Cognitive psychology can be scientific, while being interested in what goes on, unseen, inside the mind, for a number of reasons:
  - other natural sciences invoke unobservable entities and are not as a consequence rendered unscientific
  - like other sciences, cognitive psychology proceeds by modelling unobservables to produce predictions that can be tested by conducting appropriate studies
  - the advent of brain-imaging technology, though undoubtedly contentious, raises the prospect of observing processes that were previously unobservable.

5 THE COGNITIVE APPROACH

Thus far, we have talked of cognitive structures and cognitive processes. Section 3 offered some examples of historical proposals as to what kinds of things cognitive structures and processes are. Contemporary cognitive psychology equates representations with cognitive structures, and computations over these with cognitive processes.

5.1 Representation

We have emphasized the scientific nature of cognitive psychology. However, Fodor (1974) argued that psychology might be a special science – special because its subject matter, the mind, stands in a complex relation to the material, physical world – and therefore takes a different form from the natural or social sciences. Spelling out the relationship between the mind and the physical world, even between the mind and the body, is extremely difficult. Two competing intuitions have guided people's thinking about the issue. One is that the mind transcends the physical body (and the brain) – that when we say we are in love, for example, we mean more than that we are in a particular bodily or brain state. Though you may share this intuition, it is difficult indeed to say what a psychological state is if it is *not* physical. It is also difficult to reconcile this intuition with the methods of natural science – how is it possible to study something scientifically if it is not physical in nature? The competing intuition is that all aspects of humanity, including our minds, ought to be explicable as parts of the natural world, and so explicable by the natural sciences. Humans are, after all, products of natural, evolutionary pressures, shaped by the world in which we have evolved. How could we come to possess a mind that could not be explained as part of the natural, physical world?

The tension between these two intuitions is real and difficult to resolve (as you will see from Chapters 18 and 20). Here we can do no more than hint at the difficulties. One feature of the mind may go some way towards showing why the intuitions are so difficult to reconcile. It is the feature of representation.

Some things in the world have the property of being 'about' something else. Books, for example, tend to be about other things. A book on the history of Rome is about precisely that – the real events that go to make up the history of Rome. The observation is so mundane that you may never have given it a second thought. Yet this property of **aboutness** is quite extraordinary, and certainly difficult to explain within the natural sciences. A book, for example, could be described physically in terms of the arrangements of its molecules, the kinds of atoms that it comprises, its chemical compounds. We could describe its mass and volume, and measure it for electrical and magnetic properties. Yet these descriptions produce no hint as to a book's subject matter. Only when the patterns of ink are considered, not as patterns of ink, but as *words*, does it become clear what a book is about.

Few, if any, things in the natural world have this property of aboutness. It makes no sense to ask what a stone is about or what a river is about. While it makes sense to ask what a book or a newspaper is about, it makes no sense to ask what its components, the ink and paper, are about. It *does* make sense to ask what mental or cognitive processes are about – we often say to one another, 'What are you thinking about?' One way of expressing the aboutness of mental processes is to say that they involve **representations** – our thoughts *represent* possible states of affairs, our perceptions *represent* our immediate environment (generally, though not always, accurately).

The representational quality of mental processes was described by the philosopher of psychology Franz Brentano (1838–1917). Brentano believed that mental states comprise mental *acts* and mental *contents*. So, for example, my believing that Rosie, my pet cat, is lazy is a mental state – I am in the state of believing that Rosie is lazy. For Brentano, the state has a dual character: it comprises an act, corresponding to the act of believing, and a content, namely the content that Rosie is lazy. Brentano thought that mental states can differ, even if they involve the same mental act. So, for example, my believing that Rosie is lazy, and my believing that all cats are lazy, would represent two different mental states. The same act is common to both, but the beliefs are differentiated by their content: one is *about* Rosie; the other is *about* all cats.

The consequence for Brentano was that psychology needs to consider not only the internal features of the mind or brain, but also what these features are about or represent in the world. Perhaps now you can see why it is not straightforward to decide what kind of science cognitive psychology is. Whereas physics and chemistry study the material world of atoms and molecules (which do not have this representational quality), cognitive psychology studies mental states whose representational nature cannot be ignored. Consequently, cognitive psychology studies something intrinsically relational – something that spans what is in the mind and what it relates to in the world. Indeed, the issue of representation tends to distinguish the social sciences (such as sociology) from the natural sciences (like physics). Cognitive psychology, focusing on both what is represented (the world) and what does the representing (the mind), does not fall neatly into either category.

5.2 Computation

In Section 3 we considered some of the technological and theoretical antecedents to cognitive psychology. What emerged from the advances concerning theories of information and computation was the view that computers process information and provide a means for modelling and understanding the mind. As David Marr put it, 'If . . . vision is really an information processing task, then I should be able to make my computer do it . . .' (Marr, 1982, p.4).

Marr's statement hints at a deep relation between the computer and the mind. If computers process information, and information processing is what characterizes minds, then perhaps, at some deep level, the mind is computational. This claim provides a further key assumption of the cognitive approach: cognitive psychologists tend to view the mind as computational, as well as representational.

Von Eckardt (1993) suggests that there are two assumptions involved in construing the mind as computational. First is a linking assumption – the assumption that the mind is a computational device of some kind and that its capacities are computational capacities. The assumption serves to link minds (things that we wish to understand better) with computers (things that are already well understood). Second is the system assumption – this fleshes out what is meant by a computational device. Generally, the assumption tends to be that computers are systems that represent information, input, store, manipulate, and output representations, and which operate according to rules. The two assumptions work together to provide a framework for understanding the (relatively) unknown mind in terms of the known computer.

Just as with the representational assumption, the assumption that minds are computational raises many questions. One of the more pressing for cognitive psychology has been the precise form that computational models should take. This is in fact a major debate within contemporary cognitive psychology, and the issue will be referred to in one way or another in many chapters in this book (especially in Chapters 19 and 20). Broadly speaking, there have been two main proposals as to the computational models we should use to understand the mind: symbolic models and connectionist models.

5.2.1 Symbol systems

One way of understanding the idea that the mind is both representational and computational has been to suggest that the mind is a symbol system. In this view the representational qualities of the mind are expressed via the claim that the mind is symbolic and contains symbols. So, for example, my mental state that Rosie is lazy might be described as involving symbols for Rosie and laziness. The symbols together represent what the belief is about. To say that the mind is computational is to say none other than that the mind embodies (computational) mechanisms for manipulating these symbolic representations. My believing that Rosie is lazy would then involve my appropriately manipulating the symbol for Rosie and the symbol for laziness.

Newell and Simon (1976) were the first to propose that the mind is a symbol system. In their view, symbolic representations and their manipulation are the very building blocks of intelligent thought and action. Newell and Simon proposed many different properties of symbol systems but we need consider only a few. Symbol systems should comprise a basic set of symbols that can be combined to form larger symbol structures (just as the symbols for 'Rosie' and 'lazy' could be combined to form the symbolic expression 'Rosie is lazy'). Symbol systems should contain processes that operate on symbol structures to produce other symbol structures. Finally, symbol structures should represent, or be about, objects.

Newell and Simon's proposal that the mind is a symbol system amounts to the claim that the cognitive processes that underlie language, perception, memory, thinking, categorization, and problem solving will ultimately turn out to involve processes of manipulating and transforming symbolic representations. The proposal is, of course, an empirical one, and in principle the evidence could turn out either way. One way of addressing the issue is to develop models of symbol systems and compare these with empirical data (e.g. from human participants in an experiment). As you will see throughout this book, the strategy of producing computer models and comparing their performance with human data is a common one (see especially Chapter 19 for such comparisons for symbolic models). However, it is worth noting that disagreement with empirical evidence does not necessarily imply that the cognitive processes in question are not symbolic. It may well be that a different symbolic model would agree with the data much better. So, although the claim that the mind is a symbol system is empirical, it will require a considerable amount of empirical evidence to show either that the mind is symbolic or that it is not.

5.2.2 Connectionism

Cognitive psychologists have also sought to understand the mind's representational and computational qualities via an alternative framework, known as connectionism.

Connectionist models typically draw their inspiration from some of the known characteristics of the brain. So, for example, we know that neurons are highly interconnected. Seemingly they can pass information on to neurons with which they are connected either through inhibiting or enhancing the activity of those neurons. They appear to be able to process information in parallel – neurons are capable of firing concurrently. And there are many more properties besides. Connectionism describes attempts to build models of cognition out of building blocks that preserve these important properties of neural information processing. Typically, researchers simulate connectionist networks on a computer, networks that involve a number of layers of neuron-like computing units. The appeal of connectionism lies in the hope that connectionist models may ultimately stand a better chance of being successful models of cognition.

Consider the process of constructing symbolic and connectionist models in the area of language understanding, for example. A symbolic modeller might first seek to understand the representations involved in understanding language. They might posit symbolic representations of words and their meaning, of rules of grammar, and so on. They would then construct a computer program to encode the representations and manipulate them so that the program behaves sensibly. Given an input of written language, for example, the program might generate a representation of its meaning. This would be an exceptionally hard task but, were it to be successful, we could then compare the output of the program with the judgements of human language understanders to see if the program generated sensible answers.

In contrast, a connectionist modeller, though trying to represent the same kinds of information, would do this in a different way. They would seek to represent information in terms of neuron-like computing units and their interconnections. Rather than freely writing a computer program, they would seek to explain language understanding in terms of the kinds of information processing that the neuron-like units engage in. Thus connectionists seek to restrict themselves to models that have some prima facie plausibility in terms of what we know of the information-processing properties of the brain.

One of the exciting findings associated with connectionism has been that this brain-like information processing tends to produce interesting cognitive properties all on its own (some properties do not have to be explicitly programmed, unlike the case of symbolic models). For example, people tend to be good at generalizing from just a few instances – though in all likelihood you have encountered few UK Prime Ministers, if you were asked to describe the typical UK Prime Minister you could probably come up with a sensible generalization (e.g. ambitious, driven, etc.). It turns out that connectionist models tend to be able to generalize quite spontaneously, with no need for this cognitive property to be explicitly programmed.

This brief discussion aimed only to introduce these different kinds of computational model; it has of course skated over many complexities. In particular, the question as to whether the mind is better modelled as a symbol system or as a connectionist network has been and continues to be hotly debated (see, for example, Fodor and Pylyshyn, 1988; Smolensky, 1987), as you will see in Chapters 19 and 20 particularly.

SUMMARY OF SECTION 5

- Cognitive psychology is committed to the assumption that the mind is both representational and computational.
- Representations are understood as having a property of aboutness.
- Computations are understood as processes of inputting, storing, manipulating, and outputting information.
- Within cognitive psychology the mind tends to be understood in relation to either of two broad conceptions of computation:
  - computation as rule-based symbol manipulation
  - computation as neurally-inspired, as in connectionist networks.

6 LEVEL-DEPENDENT EXPLANATIONS

Linking the mind with computers raises many interesting and challenging questions. One view, commonly attributed to Marr (1982), is that cognition can be understood at (at least) three different levels.

6.1 The computational level

The first of Marr's levels (level 1) is commonly referred to as the computational level. An explanation of cognition at this level specifies *what* a computational system actually computes and *why*. The specification can be given in terms of a mapping between appropriate sets of inputs and their corresponding outputs. Consider a system that performs addition. A level 1 explanation would therefore refer to the 'plus' function, partially indicated in Table 1.2.

Marr also believed that level 1 explanations should specify why the system should compute the function that it does in order to complete a particular task. Why is it, for example, that the plus function (as opposed to

TABLE 1.2 Level 1 specification for addition. The inputs are pairs of numbers to be added and the output is their sum

Inputs	Outputs
0,0	0
0,1	1
1,0	1
2,3	5
87,123	210

multiplication) is the right function for the task of adding two numbers together?

Thus, cognitive psychologists who seek to explain some aspect of cognition at the computational level need to explain or describe the function that is computed (what the inputs and outputs are) and why that function is the appropriate one. For example, an explanation of language understanding might describe inputs that correspond to sentences or questions, and outputs that correspond to appropriate comments or responses.

6.2 The algorithmic level

Marr's level 2, commonly referred to as the algorithmic level, specifies *how* a computation is to be achieved. A level 2 explanation might describe the representations of input and output that a system employs and the algorithms that operate over these representations. For example, in computing the 'plus' function, input numbers could be represented in a number of different ways: in denary or binary notation, as Arabic or Roman numerals, or as appropriate numbers of dots. The algorithm specifies the steps involved in transforming the input representations into appropriate output representations.

To return to the example of addition, one way of representing two numbers (say, the numbers 2 and 3) involves representing them in terms of appropriate numbers of dots (i.e. •• and •••). One algorithm for adding the numbers might involve moving the two dots one at a time so that they are adjacent to the three to yield an output representation (not dissimilar to adding using an abacus). Another (formally) distinct algorithm would be to move the three dots one at a time so that they are adjacent to the two. These algorithms, and the sequence of steps they would generate, are shown in Table 1.3.

Note that these two sequences of steps achieve the same end result, 5 dots (•••••) representing the number 5. That is, although they are distinct processes, and hence distinct algorithms, at level 1 they are indistinguishable. In fact, it can be proved that there are an infinite number of different algorithms for any level 1 specification.

This obviously makes it very difficult for a cognitive psychologist to work out what algorithm to choose in order to model human performance successfully. However, there are ways of distinguishing different algorithms. For example, algorithms can bestow a considerable benefit to anyone (or anything) that deploys them. Even though a task may appear to be insoluble, or its solution appear to impose impractical demands on resources, with appropriate algorithms it may be soluble with a modicum of resources. Note how algorithm 2 in Table 1.3 completes the task in one less step than algorithm 1.

Less trivially, consider chess. One way of playing chess would be to consider all possible moves by looking ahead a certain number of steps. As one looks further ahead, however, the number of possible moves grows exponentially, and so this particular strategy would require vast amounts of memory and time. By deploying more sophisticated algorithms, ones involving heuristics and strategies that restrict the number of possible moves that need to be considered, the resource demands of the task fall rapidly. Thus, appropriate algorithms may render soluble tasks that appear

TABLE 1.3 Two algorithms and the steps they generate for computing $2 + 3$

Algorithm 1 (move one dot at a time from right to left)		Step	Algorithm 2 (move one dot at a time from left to right)	
Left	Right		Left	Right
••	•••	0	••	•••
•••	••	1	•	••••
••••	•	2		•••••
•••••		3		

insoluble, and also render them soluble within practical resource limits (Chapter 10 considers some of the stratagems of real chess experts).

To see this, consider different algorithms for multiplying 253 by 375. One option is to add 253 to itself 375 times. Another would involve adding 375 to itself 253 times. Yet another way would be to remember the products of all pairs of numbers up to, say, 400. The first and second algorithms would require a pencil and paper and a very large amount of time. By contrast, the third strategy would potentially require little time but a very large and efficient memory. A better algorithm, perhaps, would involve knowing by rote some products (say, $5 \times 200 = 1{,}000$, $3 \times 5 = 15$, etc.) and knowing that the product asked for can be decomposed as follows:

$$\begin{aligned} 253 \times 375 &= (200 + 50 + 3) \times (300 + 70 + 5) \\ &= 200 \times (300 + 70 + 5) + 50 \times (300 + 70 \\ &\quad + 5) + 3 \times (300 + 70 + 5) \\ &= (60{,}000 + 14{,}000 + 1{,}000) + (15{,}000 \\ &\quad + 3{,}500 + 250) + (900 + 210 + 15) \\ &= 75{,}000 + 18{,}750 + 1{,}125 \\ &= 94{,}875 \end{aligned}$$

Note that this algorithm involves *some* demands on memory and *some* demands on time, but doesn't place excessive demands on either.

Returning to our example of language understanding, a challenge for a cognitive psychologist would be to work out how the inputs and outputs should be represented, and algorithms for converting the former into the latter. A critical question, however, will remain: why were these particular representations and this particular algorithm chosen, and could better choices have been made?

6.3 The implementational level

Marr's level 3 is commonly referred to as the hardware or implementational level. It specifies how algorithms and representations are physically realized. In our example of addition, numbers were realized as marks on pieces of paper and movement of those marks. In a digital computer, an explanation at the implementational level would make reference to transistors, voltages, currents, diodes, and the like. If addition were implemented using an abacus, an explanation would make reference to beads sliding on rods. Were we to explain human cognitive processing in terms of Marr's level 3, then we would make reference to neurons, neurotransmitters, and so on.

Explaining cognitive processing at the implementational level presents a very real challenge. In our example of language understanding, we would have to make reference to the real neural circuits that implement language understanding, and to their actual activities whilst doing so. Though neuropsychological and neuroimaging evidence, as well as neuroscientific advances, accumulate, such an explanation exceeds the abilities of our current understanding.

6.4 Using Marr's levels

Cognitive psychologists tend to explain cognition at levels 1 and 2. That is, they pursue **functional accounts** (at level 1) and **process accounts** (at level 2). Level 3 explanations, those that refer to actual neurons, neurotransmitters, and so on, tend to be left to neuroscientists. However, there are important relations between all three levels. For example, the implementation level can constrain what counts as an appropriate algorithm. The brain may not be able to implement all algorithms, or may not implement them equally well. In a sense, connectionist models are predicated on this view – that the hardware of the brain constrains our choice of algorithm (or level 2 explanations) to those that we know the brain is good at computing. Certainly if it could be shown that a level 1 or 2 account of some cognitive phenomenon could not be implemented in neural hardware, then real doubt would be cast on the corresponding psychological explanation.

This section has focused on some of the foundational assumptions made in contemporary cognitive psychology, though very many other assumptions are also made and also tend to characterize a cognitive approach. Table 1.1 in Section 2 listed some of the more common ones and you may wish to revisit it now. You may also like to refer to this table after you have read each of the following chapters to see if you can identify which assumptions have been made and how explicitly.

SUMMARY OF SECTION 6

- Marr's levels provide a framework for understanding explanations of cognition.
- Explanations can be pitched at one of three levels:
  - computational level
  - algorithmic level
  - implementational level.
- Cognitive psychological explanations are typically expressed at levels l (functional) and 2 (process), but are assumed to be constrained by what is known about level 3.

7 CONCLUSIONS

In the previous sections we have attempted to outline some of the history of cognitive psychology, its subject matter, and also some of its core assumptions. As we have seen, cognitive psychology has a relatively long history and has made and continues to make many connections with other disciplines. To understand the nature of cognitive psychology, we have had to consider a wide range of issues, from computation to neuroimaging, from mundane but complex behaviour such as understanding language to the behaviour involved in anti-aircraft gunnery. Our survey has touched on action, perception, thinking, language, problem solving, categorization, and consciousness. We have considered the nature of scientific investigation, the importance of observation, and the need for, and practice of, sciences to posit theoretical entities that cannot be observed. We have also touched on the possibility that cognitive psychology may be a special science, perhaps somewhere between a social and a natural science.

In such a short chapter we have omitted much, and this chapter should be regarded as a partial survey of the foundations of cognitive psychology, intended to help you make the most of the chapters that follow. Most notably, we have barely touched on the different methods of cognitive psychology, though the following chapters make clear just how central these methods are to the cognitive approach.

We have not intended to suggest that cognitive psychology faces no real challenges or problems. Far from it. Most if not all of the topics we will consider in this book are still not fully understood – though cognitive psychology has proved remarkably successful so far, it remains to be seen just how well it will deliver such a full understanding. Indeed, while in topics such as attention and perception cognitive psychologists have made great progress, others, such as consciousness and emotion, still present real challenges. This is not to say that cognitive psychologists have not contributed greatly. As you will see in Chapters 16, 17, and 18

✱ ACTIVITY 1.4

In Activity 1.1, we asked you to write down what you took to be the characteristic features of a scientific study of the mind. Take a few minutes to review your list – are there some features you would want to add to the list? And are there any you would want to remove?

among others, progress has been made even though foundational questions remain.

The breadth of the many issues we have raised, as well as the results and promise of the cognitive approach that you will encounter in subsequent chapters, testify to the importance of developing a systematic and rigorous understanding of the mind. It also hints at the fascination and enjoyment that can be gained from studying cognitive psychology, something that we hope you will soon experience for yourself.

FURTHER READING

Bechtel, W. (1988) *Philosophy of Mind: An Overview for Cognitive Science*, Hillsdale, NJ, Lawrence Erlbaum Associates. Designed for cognitive scientists rather than philosophers, this is an excellent introduction to some of the more important philosophical issues that challenge any study of the mind.

Bechtel, W. and Abrahamsen, A. (2002) *Connectionism and the Mind: Parallel Processing, Dynamics, and Evolution in Networks*, 2nd edition, Oxford, Blackwell. This updated edition provides an excellent introduction to connectionism, or neural networks, the philosophical debates that surround this style of modelling, and the theoretical and technical aspects of key models.

Gardner, H. (1985) *The Mind's New Science: A History of the Cognitive Revolution*, New York, Basic Books. This book still provides an excellent introduction to and history of the key theoretical and conceptual advances that have led to our contemporary study of the mind.

REFERENCES

Bryan, W.L. and Harter, N. (1899) 'Studies on the telegraphic language: the acquisition of a hierarchy of habits', *Psychological Review*, vol.6, pp.345–75.

Chomsky, N. (1959) 'A review of B.F. Skinner's *Verbal Behavior*', *Language*, vol.35, no.1, pp.26–58.

Devlin, J.T., Russell, R.P., Davis, M.H., Price, C.J., Wilson, J., Moss, H.E., Matthews, P.M., and Tyler, L.K. (2000) 'Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task', *NeuroImage*, vol.11, pp.589–600.

Fodor, J.A. (1974) 'Special sciences', *Synthese*, vol.28, pp.77–115.

Fodor, J.A. and Pylyshyn, Z.W. (1988) 'Connectionism and cognitive architecture: a critical analysis', *Cognition*, vol.28, pp.3–71.

Greer, M.J., van Casteren, M., McLellan, S.A., Moss, H.E., Rodd, J., Rogers, T.T., and Tyler, L.K. (2001) 'The emergence of semantic categories from distributed featural representations', *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, London, Lawrence Erlbaum Associates.

Hillis, A.E. and Caramazza, A. (1991) 'Category-specific naming and comprehension impairment: a double dissociation', *Brain Language*, vol.114, pp.2081–94.

Kriegeskorte, N., Mur, M., Ruff, D., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P. (2008) 'Matching categorical object representations in inferior temporal cortex of man and monkey', *Neuron*, vol.60, no.6, pp.1126–41.

Lashley, K.S. (1951) 'The problem of serial order in behaviour', in Jeffress, L.A. (ed.) *Cerebral Mechanisms in Behaviour: The Hixon Symposium*, New York, John Wiley.

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, W.H. Freeman & Company.

Martin, A. (2007) 'The representations of object concepts in the brain', *Annual Review of Psychology*, vol.58, pp.25–45.

Miller, G.A., Galanter, E., and Pribam, K. (1960) *Plans and the Structure of Behaviour*, New York, Holt, Rinehart and Winston.

Moss, H.E., Tyler, L.K., and Taylor, K.I. (2009) 'Conceptual structure', In Gaskell, G.M. (ed.) *The Oxford Handbook of Psycholinguistics*, Oxford, Oxford University Press.

Newell, A. and Simon, H.A. (1976) 'Computer science as empirical enquiry: symbols and search', *Communications of the Association for Computing Machinery*, vol.19, pp.113–26.

Skinner, B.F. (1957) *Verbal Behavior*, New York, Appleton-Century-Crofts.

Smolensky, P. (1987) 'The constituent structure of connectionist mental states: a reply to Fodor and Pylyshyn', *The Southern Journal of Philosophy, Supplement*, vol.26, pp.137–61.

Tolman, E.C. (1932) *Purposive Behaviour in Animals and Man*, New York, Century.

Turing, A.M. (1950) 'Computing machinery and intelligence', *Mind*, vol.59, pp.433–60.

Tyler, L.K. and Moss, H.E. (2001) 'Towards a distributed account of conceptual knowledge', *Trends in Cognitive Sciences*, vol.5, no.6, pp.244–52.

Von Eckardt, B. (1993) *What Is Cognitive Science?*, Cambridge, MA, MIT Press.

Warrington, E.K. and Shallice, T. (1984) 'Category specific semantic impairments', *Brain*, vol.107, pp.829–54.

Wertheimer, M. (1912) 'Experimentelle Studien über das Sehen von Bewegung', *Zeitschrift Für Psychologie*, vol.61, pp.161–265.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

PART 1

PERCEPTUAL PROCESSES

Introduction

2 Attention

Peter Naish

3 Perception

Graham Pike, Graham Edgar, and Helen Edgar

4 Recognition

Graham Pike and Nicola Brace

INTRODUCTION

In Part 1 you will find chapters on attention, perception, and recognition. Why do we begin with these particular topics? Well, there is a fairly strong tradition of placing these topics early in books on cognition, and there are at least two reasons for this. First, there is a strong applied psychology theme to all these topics, whether it is finding better ways to present relevant information to people in safety-critical occupations, such as aircraft pilots, devising techniques for improving eye-witness identification, or designing machines that can 'see' and 'recognize'. Second, attention, perception, and recognition are all topics that concern the relationship between the mind and the world, which seems a good place to start trying to understand the mind itself. Other chapters – for example, Chapter 6 on language processing – also address the issue of how information from the world gets 'into' the mind, but the topics of attention, perception, and recognition provide particularly direct questions relating to it. Why do we become aware of some aspects of the environment rather than others? How is it that we manage to perceive those things we do become aware of? And for those things we do consciously perceive, how do we come to recognize what they are?

As you will see, these turn out to be far from simple questions and to require far from simple answers. A key issue that comes up in all three chapters has to do with distinguishing between aspects of the world (the domain of physics), how these aspects affect the body and especially the nervous system (physiology), and what mental representations result (cognitive psychology). In Chapter 2 you will learn what kinds of physical energy the auditory system uses to represent the location of a sound source; in Chapter 3 you will encounter a theory of how the visual system comes to represent Gestalt organization, which is easily mistaken for a property of the world rather than of the mind; and in Chapter 4 you will see how different aspects of the same physical face – familiarity, identity, emotional tone – are processed by different physiological pathways and have separate cognitive representations.

A further key issue that emerges in all three chapters is the fractionation of functions. It turns out that there is not just one sort of attention but many different forms of it. Similarly, it transpires that visual perception is far from being a unitary function; in fact, vision is made up of such a multitude of component processing streams that Chapter 3 has space to mention only some of them. As indicated in the previous paragraph, recognition can also be separated into different processes, and a similar fractionation will recur in later chapters in relation to other mental functions such as memory. (How we should conceptualize all these cognitive functions and their sub-components is something it might be useful to consider in the light of Chapter 5 on categorization.) Allied to the issue of how cognitive functions can be analysed into component processes are questions about which of these processes result in representations that are or are not consciously experienced, and which can be carried out in parallel, and which only one at a time.

A common theme across all the chapters is the use of neuropsychological evidence to help elucidate key issues such as those we have just identified. Injury to the brain can affect attention, perception, and recognition in quite unexpected ways. Studying the behavioural and phenomenological consequences of injury to specific parts of the brain, relating neuroanatomy to behaviour and conscious experience, throws light upon the structure of cognition by providing both tests of psychological theories and grounds from which theories may be derived.

Another issue common to all the chapters is the extent to which stored knowledge enters into the functions of attention, perception, and recognition. These functions might be purely stimulus-driven; that is, driven by physical properties of the world. But if they are not, then at what stage in processing does prior knowledge exert its influence? Do we necessarily identify a plant *before* picking it, for example? If not, why would we tend to avoid picking stinging nettles with bare hands? Do we perceive familiar faces in the same way that we perceive unfamiliar faces? If not, does familiarity also affect perception of other classes of object? It is important that answers to such questions are given within a theoretical context. When you have read the chapters, you should reflect on how well or how badly cognitive psychological theories have fared in recent decades.

In Chapter 2, Peter Naish describes such different forms of attention as attention to regions of space, attention to objects, and attention for action, but attempts finally to summarize them all under a single fairly abstract definition of the term. He shows how ideas about attention have changed and diversified over the last 50 years and considers how well the early theories have stood up to examination. He also considers the relationship between attention and consciousness, and what research on hypnosis can teach us about this. In Chapter 3, Graham Pike, Graham Edgar, and Helen Edgar consider top-down and bottom-up theories of perception, and propose a resolution in terms of perception for recognition and perception for action. They also introduce and evaluate Marr's computational framework for a bottom-up theory of perception. Lastly, in Chapter 4, Graham Pike and Nicola Brace describe and contrast two theories of object perception, as well as a model of face perception that has been implemented as a connectionist network (a type of computer model). Across the chapters you will encounter theories being tested and sometimes confirmed and sometimes found wanting. You will also meet the idea that different theories may be complements of one another rather than simply alternatives. One theory may succeed in one domain but fail in another, and vice versa for a second theory. You will also see how confidence in a theory varies with its range of application, and how confidence can be boosted if it proves possible to implement the theory as a working computer model. The challenge for the future is for theorists to develop more detailed and implementable theories of attention, perception, and recognition whilst allowing that different people may find distinct ways of doing the same thing.

This page intentionally left blank

CHAPTER 2

ATTENTION

Peter Naish

1 AUDITORY ATTENTION

For many of us the concept of attention may have rather negative connotations. At school we were told to pay attention, making us all too aware that it was not possible to listen to the teacher while at the same time being lost in more interesting thoughts. Neither does it seem possible to listen effectively to two different things at the same time. How many parents with young children would love to be able to do that! One could be excused for feeling that evolution has let us down by failing to enable us to process more than one thing at a time. If that is how you feel, then this chapter might add insult to injury because it will cite evidence that we do in fact process a good deal of the material to which we are not attending. Why, you might ask, do we go to the trouble of analysing incoming information, only to remain ignorant of the results? To attempt an answer it is necessary to consider a range of issues, stretching from registration of information by the sense organs, through the processes of perception, to the nature of awareness and consciousness. Attention is a broad and intriguing topic. That breadth makes it very difficult to offer a simple definition of the term, so I will not attempt to do so until the end of the chapter.

To cover some of this topic (we have only a chapter, and there are whole books on the subject) I shall follow an approximately historical sequence, showing how generations of psychologists have tackled the issues and gradually refined and developed their theories. You will discover that initially there seemed to them to be only one role for attention, but that gradually it has been implicated in an ever-widening range of mental processes. As we work through the subject, two basic issues will emerge. One is concerned with the mechanisms of attention and raises questions such as:

- How much material can we take in at once?
- What happens to information to which we did not attend?
- In what circumstances does attention fail, allowing unwanted information to influence or distract us?

The other theme has more to do with the purpose behind the mechanisms and raises questions concerning why we experience the apparent limitations of attention:

- Are the limitations simply an inevitable characteristic of a finite brain?
- Have we evolved to exhibit attention – that is, does it confer advantages?

We shall begin to explore these issues by looking at the ways in which one of our senses (hearing) has developed to facilitate attention.

1.1 Disentangling sounds

If you are still feeling aggrieved about the shortcomings of evolution, then you might take heart from the remarkable way in which the auditory system has evolved so as to avoid a serious potential problem. Unlike our eyes, our ears cannot be directed so as to avoid registering material that we wish to ignore; whatever sounds are present in the environment, we must inevitably be exposed to them. In a busy setting such as a party we are swamped by simultaneous sounds – people in different parts of the room all talking at the same time. Indeed, our ability to listen to just one out of a medley of sounds is known as the **cocktail party effect**. An analogous situation for the visual system would be if several people wrote superimposed messages on the same piece of paper and we then attempted to pick out one of the messages and read it. Because that kind of visual superimposition does not normally occur, there have been no evolutionary pressures for the visual system to find a solution to the problem (though see below). The situation is different with hearing, but the possession of two ears has provided the basis for a solution.

Figure 2.1 shows a plot of sound waves recorded from inside a listener's ears. You can think of the up and down movements of the wavy lines as representing the in and out vibrations of the listener's ear drums. The sound was of a single hand clap, taking place to the front left of the listener. You will notice that the wave for the right ear (i.e. the one further from the sound) comes slightly later than the left (shown by the plot being shifted to the right). This right-ear plot also goes up and down far less, indicating that it was less intense, or in hearing terms that it sounded less loud at that ear. These differences in timing and intensity are important to the auditory system, as will be explained.

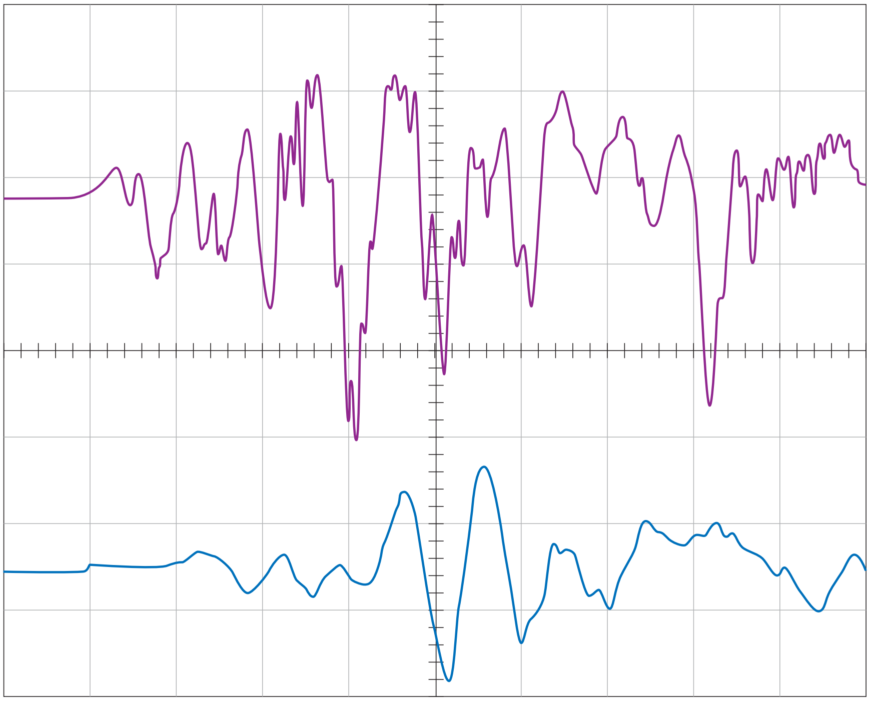


FIGURE 2.1 The waveform of a hand clap, recorded at the left (upper trace) and right (lower trace) ears. Horizontal squares represent durations of 500 microseconds (a microsecond is one-millionth of a second); vertical divisions are an arbitrary measure of sound intensity.

Figure 2.2(a) represents sound waves spreading out from a source and passing a listener's head. Sound waves spread through the air in a very similar way to the waves (ripples) spreading across a pond when a stone is thrown in. For ease of drawing, the figure just indicates a 'snapshot' of the positions of the wave crests at a particular moment in time. Two effects are shown. First, the ear further from the sound is slightly shadowed by the head, so receives a somewhat quieter sound (as in Figure 2.1). The head is not a very large obstacle, so the intensity difference between the ears is not great; however, the difference is sufficient for the auditory system to register and use it. If the sound source were straight ahead there would be no difference, so the size of the disparity gives an indication of the sound direction. The figure also shows a second difference between the ears: a different wave part (crest) has reached the nearer left ear compared with the right (which is positioned somewhere in a trough between two peaks). Once again, the inter-aural difference is eliminated for sounds coming from straight ahead, so the size of this difference also indicates direction.

Why should we make use of both intensity and wave-position differences? The reason is that neither alone is effective for all sounds. I mentioned that the head is not a very large obstacle; what really counts is how large it is compared with a **wavelength**. The wavelength is the distance from one wave crest to the next. Sounds that we perceive as low-pitched have long wavelengths – longer in fact than the width of the head. As a result, the waves pass by almost as if the head was not there. This means that there is negligible intensity shadowing, so the intensity cue is not available for direction judgement with low-pitched sounds. In contrast, sounds that we experience as high pitched (e.g. the jingling of coins) have wavelengths that are shorter than head width. For these waves the head is a significant obstacle and shadowing results. To summarize, intensity cues are available only for sounds of short wavelength.

In contrast to the shadowing effect, detecting that the two ears are at different positions on the wave works well for long wavelength sounds. However, it produces ambiguities for shorter waves. The reason is that if the wave crests were closer than the distance from ear to ear, the system would not be able to judge whether additional waves should be allowed for. Figure 2.2(b) shows an extreme example of the problem. The two ears are actually detecting identical parts of the wave, a situation that is normally interpreted as indicating sound coming from the front. As can be seen, this wave actually comes from the side. Our auditory system has evolved so that this inter-ear comparison is

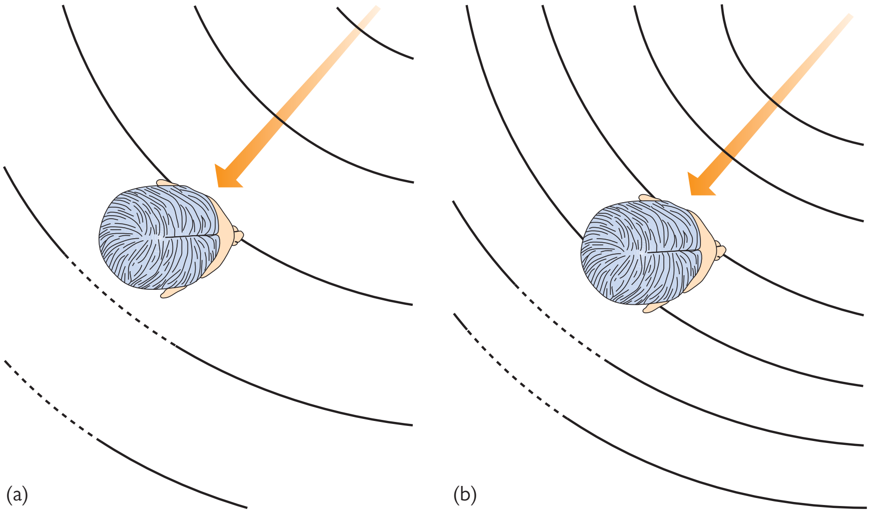


FIGURE 2.2 Curved lines represent wave crests of a sound approaching from a listener's front left. In (b) the sound has a shorter wavelength (higher pitched) than in (a), so waves are closer together with a crest at each ear.

made only for waves that are longer than the head width, so the possibility of the above error occurring is eliminated. Consequently, this method of direction finding is effective only for sounds with long wavelengths, such as deeper speech sounds.

You will notice that the two locating processes complement each other perfectly, with the change from one to the other taking place where wavelengths match head width. Naturally occurring sounds usually contain a whole range of wavelengths, so both direction-sensing systems come into play and we are quite good at judging where a sound is coming from. However, if the only wavelengths present are about head size, then neither process is fully effective and we become poor at sensing the direction. Interestingly, animals have evolved to exploit this weakness. For example, pheasant chicks (that live on the ground and cannot fly to escape predators) emit chirps that are in the 'difficult' wavelength range for the auditory system of a fox. The chicks' mother, with her bird-sized head, does not have any problems at the chirp wavelength, so can find her offspring easily. Before the advent of personalized ring tones, mobile telephone manufacturers seemed to have followed the same principle. To my ears they appeared to be using ring tones with frequencies that made it impossible to know whether it was one's own or someone else's phone that was ringing!

1.2 Attending to sounds

From the above, you will appreciate that the auditory system is able to separate different, superimposed sounds on the basis of their different source directions. This makes it possible to attend to any one sound without confusion, and we have the sensation of moving our 'listening attention' to focus on the desired sound. For example, as I write this I can listen to the quiet hum of the computer in front of me, or swing my attention to the bird song outside the window to my right. Making that change feels almost like swinging my eyes from the computer to the window, and the term **spotlight of attention** has been used to describe the way in which we can bring our attention to bear on a desired part of the environment.

✱ ACTIVITY 2.1

1 Set up a sound source (the radio, say) and then listen to it from across the room. Turn sideways-on so that one ear faces the source. Now place a finger in that nearer ear, so that you can hear the sound only via the more distant ear. You should find that the sound seems more muffled and deeper, as if someone had turned down the treble on the tone control. This occurs because the shorter wavelength (higher pitched) sounds cannot get round your head to the uncovered ear. In fact you may still hear a little of those sounds, because they can reflect from the walls and so reach your uncovered ear 'the long way round'. Most rooms have sufficient furnishings (carpets, curtains, etc.) to reduce these reflections, so you probably will not hear much of the higher sounds. However, if you are able to find a rather bare room (bathrooms often have hard, shiny surfaces) you can use it to experience the next effect.

2 Do the same as before, but this time you do not need to be sideways-on to the sound. If you compare your experiences with and without the finger in one ear you will probably notice that, when you have the obstruction, the sound is more 'boomy' and unclear. This lack of clarity results from the main sound, which comes directly from the source, being partly smothered by slightly later echoes, which take longer routes to your ear via many different paths involving reflections off the walls, etc. These echoes are still there when both ears are uncovered, but with two ears your auditory system is able to detect that the echoes are coming from different directions from the main sound source, enabling you to ignore them. People with hearing impairment are sometimes unable to use inter-aural differences, so find noisy or echoing surroundings difficult.

My account so far has explained the mechanisms that stop sounds becoming 'jumbled' and reminds us that, subjectively, we listen to just one of the disentangled sounds. It seems obvious that they would need disentangling to become intelligible, but why do we then attend to only one? That question leads us into the early history of attention research.

One of the first modern researchers to investigate formally the nature of auditory attention was Broadbent (1952, 1954), who used an experimental technique known as **dichotic listening**. This offers a way of presenting listeners with a simplified, more easily manipulated version of the real world of multiple sounds. Participants wear a pair of headphones and receive a different sound in each ear; in many studies the sounds are recorded speech, with each ear receiving a different message. Broadbent and others (e.g. Treisman, 1960) showed that, after attending to the message in one ear, a participant could remember virtually nothing of the unattended message that had been played to the other ear, often not even whether it had been spoken in a foreign language.

Broadbent's experiments showed that the last statement needed some refinement. If the two messages were very short, say just three words in each ear, then the participant could report what had been heard by the unattended ear. Only when the attended message lasted more than a few seconds was the material in the other ear lost. The system behaved as if there were a short-lived store that could hold a small segment of the unattended material until analysis of the attended words was complete. The store's quality of hanging on to a sound for a short time, like a dying echo, led to it being termed the **echoic memory**.

It was also shown that people would often be aware of whether an unattended voice had been male or female, and they could use that distinction to follow a message. Two sequences of words were recorded, one set by a woman and the other by a man. Instead of playing one of these voice sequences to each headphone, the words were made to alternate. Thus, the man's voice jumped back and forth, left to right to left, while the woman's switched right to left to right. In this situation participants were able to abandon the normal 'attending by ear' procedure, and instead report what a particular speaker had said; instead of using location as a cue for attention, they were using the pitch of the voice.

The explanation for these findings seemed straightforward. Clearly the brain had to process the information in a sound in order to understand it as speech. In this respect, the brain was rather like a computer

BOX 2.1 RESEARCH STUDY Application of research on auditory attention

Donald Broadbent's early career included research for the UK Ministry of Defence, and his findings often led to innovation. One problem he addressed was the difficulty pilots experienced when trying to pick out a radio message from a number of interfering stations (radio was less sophisticated then). Pilots' headphones delivered the same signals to each ear, so it was not possible to use inter-aural differences to direct attention to the wanted message. Broadbent devised a stereo system, which played the desired signal through both headphones, while the interference went only to one or the other. This made the interference seem to come from the sides, while the signal sounded as if it was in the middle (identical waves at the two ears). In effect, this was dichotic listening, with a third (wanted) signal between the other two. The improvement in intelligibility was dramatic, but when Broadbent played a recording to officials they decided that it was so good that he must have 'doctored' the signal! The system was not adopted. Decades later, I demonstrated (Naish, 1990) that using stereo and giving a directional quality to the headphone warning sounds used in aircraft cockpits could result in significantly shorter response times. Thus, the warning indicating an approaching missile could be made to seem as if it were coming from the missile direction, so speeding the pilot's evasive measures. Fighter aircraft may at last incorporate '3D' sound.

processing information (computers were beginning to appear at that time), and everyone knew that computers could only process one thing at a time – that is, serially. Obviously (theorists thought) the brain must be serial too, so while processing the information of interest, it needed to be protected from all the rest: it needed to attend and select. However, the earliest stages of processing would have to take place in parallel (i.e. taking in everything simultaneously), ensuring that all information would potentially be available. These initial processes would have to utilize selection procedures based upon very simple characteristics of the signals, because anything more complex would demand serial processing and that couldn't start until *after* the selection. The selection processes identified were indeed simple: attention was directed either on the basis of the direction of a sound, or on whether it was higher or lower pitched. Broadbent's (1954) theory was that, after the first early stage of parallel information capture, a 'gate' was opened to one stream of information and closed to the rest.

1.3 Eavesdropping on the unattended message

It was not long before researchers devised more complex ways of testing Broadbent's theory of attention, and it soon became clear that it could not be entirely correct. Even in the absence of formal experiments, common experiences might lead one to question the theory. An oft-cited example is the own name effect. Imagine you are attending a noisy party but your auditory location system is working wonderfully, enabling you to focus upon one particular conversation. Suddenly, from elsewhere in the room, you hear someone mention your name! If you were previously selecting the first conversation, on the basis of its direction and the speaker's voice, then how did your 'serial' brain manage to process another set of sounds in order to recognize your name?

Addressing this puzzle, Treisman (1960) suggested that, rather than the all-or- nothing selection process implied by Broadbent, the ability to pick out one's name could be explained by an attenuation process. To attenuate means to reduce the intensity, so the process would function as if there were a filter, 'turning the volume down' for all but the attended signal. Although that would leave most unattended material so attenuated as to be unnoticed, for a signal to which we were very sensitive, such as our own name, there would be sufficient residual information for it to be processed and hence attract our attention. Treisman devised a series of ingenious experiments that supported this idea. Many of her studies involved shadowing, a dichotic listening technique that requires the participant to repeat aloud everything that is heard in one ear, following like a shadow close behind the spoken message. (NB This is not to be confused with the very different 'head shadowing' referred to earlier.) This task demands full concentration on the shadowed message, so when it ceases the participant appears to be completely ignorant of what was said in the other ear.

In one experiment Treisman actually made the storylines in the messages swap ears in the middle of what was being said. Thus, the left ear might hear:

> *Little Red Riding Hood finally reached the cottage, but the wicked wolf was in * beds; one was large, one medium, and one small.*

Meanwhile, the right ear would receive:

> *When she had finished the porridge, Goldilocks went upstairs and found three * bed, dressed in the grandmother's clothes.*

The asterisks indicate where the storylines swap ears. The interesting finding is that when asked to shadow one ear participants tend to end by shadowing the other, because they follow the sense of the story. Broadbent's position could not explain that, since the listener could not know that the story continued in the other ear, if that ear had been completely ignored. Treisman, on the other hand, claimed that the story temporarily sensitized the listener to the next expected words, just as with the permanent sensitization associated with our own name. Sensitization of this temporary kind is known as priming, and many experimental techniques have demonstrated its existence. For example, in a lexical decision task (a task that requires participants to indicate as quickly as possible whether or not a string of letters spells a real word) people can respond much more quickly to a word if it is preceded by another related to it. For example, the 'Yes' is given to *doctor* (yes, because it is a word) more quickly when presented after the word *nurse* than when following the word *cook*.

Treisman's ideas stimulated a succession of experiments, some seeming to show that information could 'get through' from a wider range of stimuli than one's own name or a highly predictable word in a sentence. For example, Corteen and Wood (1972) carried out a two-part experiment. Initially they presented their participants with a series of words, and each time a word from a particular category (city name) appeared the participant was given a mild electric shock. In this way, an association was formed between the shock and the category. Although the shocks were not really painful, they inevitably resulted in something like apprehension when one of the critical words was presented. This response (which once learned did not require the shocks in order for it to continue for a while) could be detected as a momentary change in skin electrical resistance. The sweat glands of a nervous person begin to secrete and the salty fluid is a good conductor of electricity, so it lowers the skin's resistance to a small (non-shocking) current. The change is known as the **galvanic skin response (GSR)** and has been used in so-called lie detectors; a person trying to hide the truth may be expected to be nervous, hence they produce the electrical effect. Corteen and Wood connected their participants to GSR apparatus when they started the second part of the experiment: a dichotic listening task. As usual, participants could later remember nothing about the unattended message, but the GSR showed that each time the ignored ear received one of the 'shocked' words there was a response. Moreover, a GSR was detected even to words of the same category that had not been presented during the shock-association phase. This generalizing of the response to un-presented words strengthens the claim that their meanings were established, even when not consciously perceived.

Not surprisingly, at this stage of research into auditory attention a number of psychologists began to question the idea that the brain could not process more than one signal at a time. Deutsch and Deutsch (1963) suggested that *all* messages received the same processing whether they were attended or not; Norman (1968) proposed that unattended information must at least receive sufficient processing to activate relevant semantic memories (i.e. the memory system that stores the meanings of words; see Chapter 8). These suggestions certainly explained the intriguing dichotic listening results, showing people to be influenced by material of which they seemed to have no knowledge. However, the ideas, if true, would require the brain to be far more parallel in its function than had been supposed. At that time there was neither an analogue by which parallel processing could be conceptualized, nor sufficient neuroanatomical information to contribute to the debate. Today there is ample evidence of the parallel nature of much of the brain's processing and, additionally, computers have advanced to the stage where brain-like parallel processing can be emulated (see Chapter 19). Thus, modern researchers have no difficulty in conceptualizing parallel processing and the nature of the attention debate has shifted somewhat. Nevertheless, recent studies have also revealed that early stages of analysis are modified by attention, effects that Broadbent would have immediately recognized as examples of filtering. We shall explore these issues in more depth, after first considering the nature of attention in visual processing.

SUMMARY OF SECTION 1

The auditory system is able to process sounds in such a way that, although several may be present simultaneously, it is possible to focus upon the message of interest. However, in experiments on auditory attention there have been contradictory results concerning the fate of the unattended material:

- The auditory system processes mixed sounds in such a way that it is possible to focus upon a single wanted message.

- Unattended material appears not to be processed:
  - The listener is normally unable to report significant details concerning the unattended information.
  - Only the most recent unattended material is available, while still preserved in the echoic memory.
- These results suggest parallel acquisition of all available information, followed by serial processing to determine meaning for just one attended message.
- Although there is little conscious awareness of unattended material, it may receive more processing than the above results imply:
  - Words presented to the unattended ear can produce priming and physiological effects.
  - Participants trying to 'shadow' one ear will follow the message to the other ear.
- These results imply that processing takes place in parallel, to the extent that meaning is extracted even from unattended material.

2 VISUAL ATTENTION

I introduced Section 1 by suggesting that the auditory system had a special problem: unlike the visual system, it needed processes that would permit a listener to attend to a specific set of sounds without being confused by the overlap of other, irrelevant noises. The implication of that line of argument was that vision had no need of any such system. However, although we do not see simultaneously *everything* that surrounds us, we can certainly see more than one thing at a time. Earlier, I wrote of attending to the sound of the computer in front of me, or of the birds to one side. I can do much the same visually. While keeping my eyes directed to the computer screen, I can either attend to the text I am typing or, out of the corner of my eye, I can be aware of the window and detect a bird when it flies past. If our eyes can receive a wide range of information in parallel, does that give the brain an attentional problem analogous to that of disentangling sounds? If visual information is handled in much the same way as auditory information seems to be, then we might expect the various items in the field of view to activate representations in memory simultaneously. That should lead to effects equivalent to those found in listening experiments; in other words, it might be possible to show that we are influenced by items that we did not know we had seen. We shall examine evidence of this shortly, but I shall first draw your attention to another area of similarity between hearing and seeing.

I pointed out at the start of Section 1.1 that, whereas we often have to follow one speech stream while ignoring others, we do not normally have to disentangle overlapping handwriting. However, it is worth bearing in mind that visual objects do overlap and hide parts of each other, and the brain certainly has the problem of establishing which components of the image on the retina 'go together' to form an object. This issue is examined in more depth in Chapter 3.

As with hearing, a variety of cues are available to help in directing visual attention. Taking my window again as an example, I can either look at the glass and see a smear (I really must get round to washing the window!) or I can look through that to the magpie sitting chattering in the apple tree. In this kind of situation we use distance to help separate objects, in much the same way as we use direction in hearing. However, we can deploy our attention in a more sophisticated way than simply on the basis of distance, as can be demonstrated by another aircraft-related example.

Military jets are often flown very fast and close to the ground (to avoid radar detection), requiring the pilot to attend intently to the outside view. At the same time, there are various pieces of information, traditionally displayed on instruments within the cockpit, which the pilot must check frequently. To avoid the pilot having to look down into the cockpit, the **head-up display (HUD)** was developed. This comprises a piece of glass, just in front of the pilot, in which all the vital information is reflected. The pilot can read the reflection, or look through it to the outside world, just as one can look at reflections in a shop window, or look through to the goods on display. With a simple reflection, the pilot would still have to change focus, like me looking at the smear or the bird. However, modern HUDs use an optical system that makes the information reflected in the display appear to be as far away as the outside scene. This saves valuable re-focusing time. Nevertheless, although the numerals in the HUD now appear to be located at the same distance as, say, a runway, pilots still have the sensation of focusing on one or the other; if they are reading their altitude they are relatively unaware of the scene on which it is superimposed. This suggests (as we shall see in more detail later) that visual attention can be linked to specific objects rather than to general regions of space, very much as auditory attention can follow a particular speaker's voice or the sense of a sentence.

2.1 Knowing about unseen information

An obvious difference between hearing and seeing is that the former is extended in time, while the latter extends over space. So, for example, we can listen to a spoken sentence coming from one place, but it takes some time to hear it all. In contrast, a written sentence is spread over an area (of paper, say) but, as long as it is reasonably short, the whole thing can be seen almost instantly. Nevertheless, seeing does require some finite time to capture and analyse the information. This process can be explored by presenting letters or words for a short, measured period of time; nowadays they are shown on a computer screen, but early research used a dedicated piece of apparatus called a **tachistoscope**. Just how long was required to register a small amount of information was investigated by Sperling (1960), who showed participants grids of letters, arranged as three rows of four letters each. If such a display was presented for 50 ms (i.e. 50 milliseconds, which is one twentieth of a second), people were typically able to report three or four of the letters; the rest seemed to have remained unregistered in that brief period of time.

Sperling explored this further. He cued participants with a tone, indicating which of the three rows of letters they should try to report; a high note for the top row, lower for middle, and deep for bottom. Crucially, the tones were not presented until just *after* the display had disappeared, meaning that participants were not able to shift their attention in preparation for the relevant row of letters when presented; it had already been presented. Strange as it seemed, people were still able to report three or four items from the cued row. Since they did not know until after the display had gone which row would be cued, this result implied that they must have already registered most of the letters in *every row*; in other words, between 9 and 12 letters in total. This apparent paradox, of seeming to know about a larger proportion of the items when asked only to report on some of them, is called the **partial report superiority effect**. The effect was also observed if letters were printed with six in red and six in black ink, then two tones used to indicate which colour to report. Participants seemed to know as much about one half (the red, say) as they did about all 12, implying that, although they could not report all the letters, there was a brief moment when they did have access to the full set and could choose where to direct their attention. The 'brief moment' was equivalent to the echoic memory associated with dichotic listening experiments, so the visual counterpart was termed an **iconic memory** (an icon being an image). All the material seemed to be captured in parallel, and for a short time was held in iconic memory. Some was selected for further, serial processing on the basis of position or colour, these being analogous to position and voice pitch in dichotic listening tasks. Unselected material (the remaining letters) could not be remembered.

With the close parallels between these auditory and visual experiments, you will not be surprised to learn that the simple selection and serial processing story was again soon challenged, and in very similar ways. Where the hearing research used shadowing to prevent conscious processing of material, the visual experiments used **backward masking**. Masking is a procedure in which one stimulus (the target) is rendered undetectable by the presentation of another (the mask); in backward masking the mask is presented after the target, usually appearing in the order of 10–50 ms after the target first appeared. The time between the onset of the target display and the onset of the mask is called the **stimulus onset asynchrony (SOA)**. The target might be an array of letters or words; this disappears after a few tens of milliseconds to be replaced by the mask, which is often a random pattern of lines. The SOA can be adjusted until participants report that they do not even know whether there has been a target, let alone what it was. In such circumstances the influence of the masked material seems sometimes still to be detected via priming effects. Thus, Evett and Humphreys (1981) used stimulus sequences containing two words, both of which were masked. The first was supposed to be impossible to see, while the second was very difficult. It was found that when the second word was related to the first (e.g. 'tiger' following 'lion') it was more likely to be reported accurately; the first, 'invisible' word apparently acted as a prime.

Claims such as these have not gone unchallenged. For example, Cheesman and Merikle (1984) pointed out that although participants say they cannot see masked words, they often do better than chance when forced to guess whether or not one had actually been presented. These researchers insisted that proper conclusions about extracting meaning from unseen material could be made only if the material was truly unseen; that is, when the participants could do no better than chance at guessing whether there had been a word displayed. Under these conditions they found no evidence for priming by masked words. However, more recently researchers have provided persuasive evidence that meaning *can* be extracted from material of which the participant is unaware. This is worth examining in more detail.

Pecher *et al.* (2002) used the Evett and Humphreys (1981) technique, but with modifications. As in the earlier study, they showed a potential prime (e.g. 'lion'), followed by a hard-to-see masked target (e.g. 'tiger'). However, there were two changes in this study. First, the priming word could be displayed either for a very short time, so that it was allegedly undetectable, or it was shown for a duration of 1 second, giving ample time for reading and guaranteeing a priming effect.

The second change was to use two sets of trials. In one, the following target was almost always (90 per cent of the time) related to the prime (e.g. 'lion' followed by 'tiger', 'bread' then 'butter', etc.). In the other set of trials only 10 per cent of trials used related words. For remaining trials the stimuli were unrelated, so that the first word was not strictly a prime (e.g. 'list' followed by 'tiger'). The results of this study are summarized in Table 2.1.

The effects are best appreciated by looking first at the final two columns of figures, showing the results when the first word was displayed for 1 second. For the condition where only 10 per cent of targets were related to the preceding word, 70 per cent of those targets were correctly identified when there was a relationship. The hit rate fell to 55 per cent when the targets were not related, so the priming effect produced a 15 percentage point advantage (70 – 55 = 15). The last column shows a massive 91 per cent hit rate for related words when there was a 90 per cent chance that they would be related to the preceding prime. The

TABLE 2.1 The percentage of targets correctly reported under various priming conditions

	Short duration prime		1 second prime	
	10% related	90% related	10% related	90% related
Related words	56	52	70	91
Unrelated words	49	43	55	51
Priming advantage	7	9	15	40

Source: adapted from Pecher et al., 2002 priming advantage in this condition has risen to 40 points. Why does the benefit of a related prime jump from 15 to 40 when the targets are more likely to be related to the primes? The answer is that, when there is a high chance that they will be related, participants spot the connection and try to guess what the target must have been: they often guess correctly. Notice that they can do this only because the prime word was clearly visible. Look now at the corresponding figures for when the prime was displayed very briefly. Here the priming advantages (7 and 9 percentage points) are far more modest (but statistically significant). However, the important result is that the change from 10 per cent to 90 per cent relatedness does not produce the large increase in the priming effect observed in the 1 second condition. The small increase from 7 per cent to 9 per cent was not statistically significant. It can be concluded that participants were unable to guess in the brief condition, so presumably had not been able to identify the prime words. Nevertheless, those words did produce a small priming effect, so they must have received sufficient analysis to activate their meaning.

2.2 Towards a theory of parallel processing

When people are asked to guess about masked material, they are commonly able to provide some information, but it often lacks detail. For example, if participants in a Sperling-type experiment have recalled three letters but are pressed for more, then they can often provide one or two. However, they generally do not know information such as whereabouts in the display the letters occurred or what colour they were. These, of course, are exactly the kinds of details that can be used to select items for report, and were believed to be usable in that role because they were characteristics that could be processed quickly and in parallel. The guessing results seem to turn the logic on its head because the presumed complex information, such as letter identities, is discovered, while the simple colour and position information is unavailable. Coltheart (1980) offered an elegant solution to this problem, built around the semantic/episodic distinction used when describing memory (see Chapter 8). In the context of letters, **semantic** information would be the basic knowledge of letter identity. **Episodic** detail links the general identity to a specific occurrence; detail such as the fact that 'N' is in upper-case type, is printed in red, and appears at the start of the third row of letters. Coltheart proposed that items do not normally reach conscious awareness unless both the semantic and episodic detail are detected. So, for example, one would not expect to be having an 'N-feeling' (semantic) if there were not a letter with some specific characteristics (size, colour, etc.) in the field of view!

It has become clear from electrophysiological studies that visual item identification occurs in a different region of the cortex from the areas that respond to colour or location. These different kinds of information have to be united, and this process, Coltheart (1980) suggests, takes time and attention. According to this account, Sperling's 12 letters, or even Evett and Humphrey's *lion*, are indeed processed in parallel to cause semantic activation, but the viewer will not become aware of this unless they are able to assign the corresponding episodic details. Nevertheless, if pressed, the participant may sometimes admit to 'having a feeling' that an item might have been presented, although they do not know what it looked like (see also Chapter 8 for a discussion of the semantic–episodic distinction).

The important point to note in the above account is that attention is no longer being described as the process that selects material for complex serial processing (e.g. word identification). Instead, Coltheart suggests that attention is required to join the products of two parallel processes: the identification and the episodic characterization. This idea, that attention is concerned with uniting the components of a stimulus, is not unlike a theory that Treisman has developed (after her early auditory attention work she turned to visual attentive processes). We shall shortly consider Treisman's work (which does not involve backward masking) but first we should look a little further at what masking actually does to the processing of a stimulus.

2.3 Rapid serial visual presentation

It has been known for a long time that backward masking can act in one of two ways: **integration** and **interruption** (Turvey, 1973). When the SOA between target and mask is very short, integration occurs; that is, the two items are perceived as one, with the result that the target is difficult to report, just as when one word is written over another. Of more interest is masking by interruption, which is the type we have been considering in the previous section. It occurs at longer SOAs, and interruption masking will be experienced even if the target is presented to one eye and the mask to the other. This **dichoptic** (two-eyed) interaction must take place after information from the two eyes has been combined in the brain; it could not occur at earlier stages. In contrast, integration masking does not occur dichoptically when target and mask are presented to separate eyes, so presumably occurs quite early in analysis, perhaps even on the retina. On this basis, Turvey (1973) described integration as peripheral masking and interruption as central masking, meaning that interruption occurred at a level where more complex information extraction was taking place.

Another early researcher in the field (Kolers, 1968) described the effect of a central (interruption) mask by analogy with the 'processing' of a customer in a shop. If the customer (equivalent to the target) comes into the shop alone, then s/he can be fully processed, even to the extent of discussing the weather and asking about family and holidays. However, if a second customer (i.e. a mask) follows the first, then the shopkeeper has to cease the pleasantries and never learns about the personal information. The analogy was never taken further, and of course it is unwise to push an analogy too far. Nevertheless, one is tempted to point out that the second customer is still kept waiting for a while. Where does that thought take us? It became possible to investigate the fate of following stimuli, in fact whole queues of stimuli, with the development of a procedure popularized by Broadbent (Broadbent and Broadbent, 1987), who, like Treisman, had moved on from auditory research. The procedure was termed Rapid Serial Visual Presentation, in part, one suspects, because that provided the familiar abbreviation RSVP; participants were indeed asked to *répondez s'il vous plaît* with reports of what they had seen.

Unlike the traditional two-stimulus, target/mask pairing, **rapid serial visual presentation (RSVP)** displayed a series of stimuli in rapid succession, so each served as a backward mask for the preceding item. SOAs were such that a few items could be reported, but with difficulty. Typical timings would display each item for 100 ms with a 20 ms gap between them; the sequence might contain as many as 20 items. Under these conditions stimuli are difficult to identify and participants are certainly unable to list all 20; they are usually asked to look out for just two. In one variation, every item except one is a single black letter. The odd item is a white letter and this is the first target; the participant has to say at the end of the sequence what the white letter was. One or more items later in the sequence (i.e. after the white target), one of the remaining black letters may be an 'X'. As well as naming the white letter, the participant has to say whether or not X was present in the list. These two targets (white letter and black X) are commonly designated as T1 and T2. Notice that the participant has two slightly different tasks; for T1 (which will certainly be shown) an unknown letter has to be identified, whereas for T2 the task is simply to say whether a previously designated letter was presented. These details, together with a graph of typical results, are shown in Figure 2.3.

As can be seen from the graph in Figure 2.3(b), T2 (the X) might be spotted if it is the item immediately following T1, but thereafter it is less likely that it will be detected unless five or six items separate the two. What happens when it is not detected? As you may be coming to expect, the fact that participants do not report T2 does not mean that they have not carried out any semantic analysis upon it. Vogel *et al.* (1998) conducted an RSVP experiment that used words rather than single letters. Additionally, before a sequence of stimuli was presented, a clear 'context' word was displayed for a comfortable 1 second. For example, the context word might be *shoe* and then the item at T2 could be *foot*. However, on some presentations T2 was not in context; for example, *rope*. While

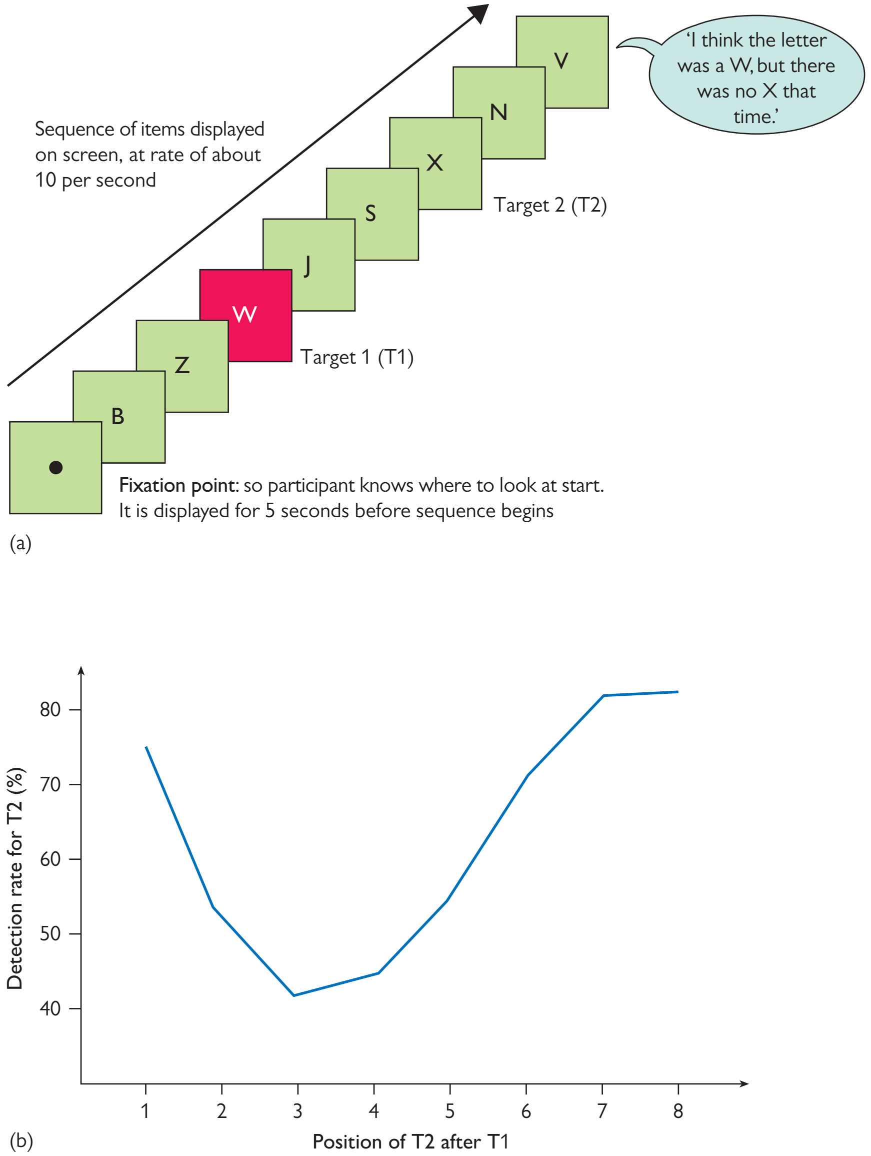


FIGURE 2.3 The RSVP technique: (a) The sequence of stimuli, shown in the same location on a computer screen, in which the participant has to identify a white letter and then decide whether an X was also present; (b) Typical results, showing the likelihood of detecting the X when presented in the first and subsequent positions following the white target.

participants were attempting to report these items, they were also being monitored using EEG (electroencephalography). The pattern of electrical activity measured via scalp electrodes is known to produce a characteristic 'signature' when what might be called a mismatch is encountered. For example, if a participant reads the sentence *He went to the café and asked for a cup of tin*, the signature appears when *tin* is reached. The Vogel *et al.* (1998) participants produced just such an effect with sequences such as *shoe – rope*, even when they were unable to report seeing *rope*. This sounds rather like some of the material discussed earlier, where backward masking prevented conscious awareness of material that had clearly been detected. However, the target in the RSVP situation appears to be affected by something that happened *earlier* (i.e. T1), rather than by a following mask. The difference needs exploring and explaining.

Presumably something is happening as a result of processing the first target (T1) that temporarily makes awareness of the second (T2) very difficult. Measurements show that for about 500 to 700 ms following T1, detection of T2 is lower than usual. It is as if the system requires time to become prepared to process something fresh, a gap that is sometimes known as a **refractory period**, but that in this context is more often called the **attentional blink**, abbreviated to AB. While the system is 'blinking' it is unable to attend to new information.

Time turns out not to be the only factor in observing an AB effect ('AB effect' will be used as a shorthand way of referring to the difficulty of reporting T2). Raymond *et al.* (1992) used a typical sequence of RSVP stimuli but omitted the item immediately following the *first* target. In other words, there was a 100 ms gap rather than another item following. Effectively, this meant that the degree of backward masking was reduced, and not surprisingly this resulted in some improvement in the report rate for T1. Very surprisingly, it also produced a considerable improvement in the reporting of T2; the AB effect had vanished (see Figure 2.4(a)). How did removing the

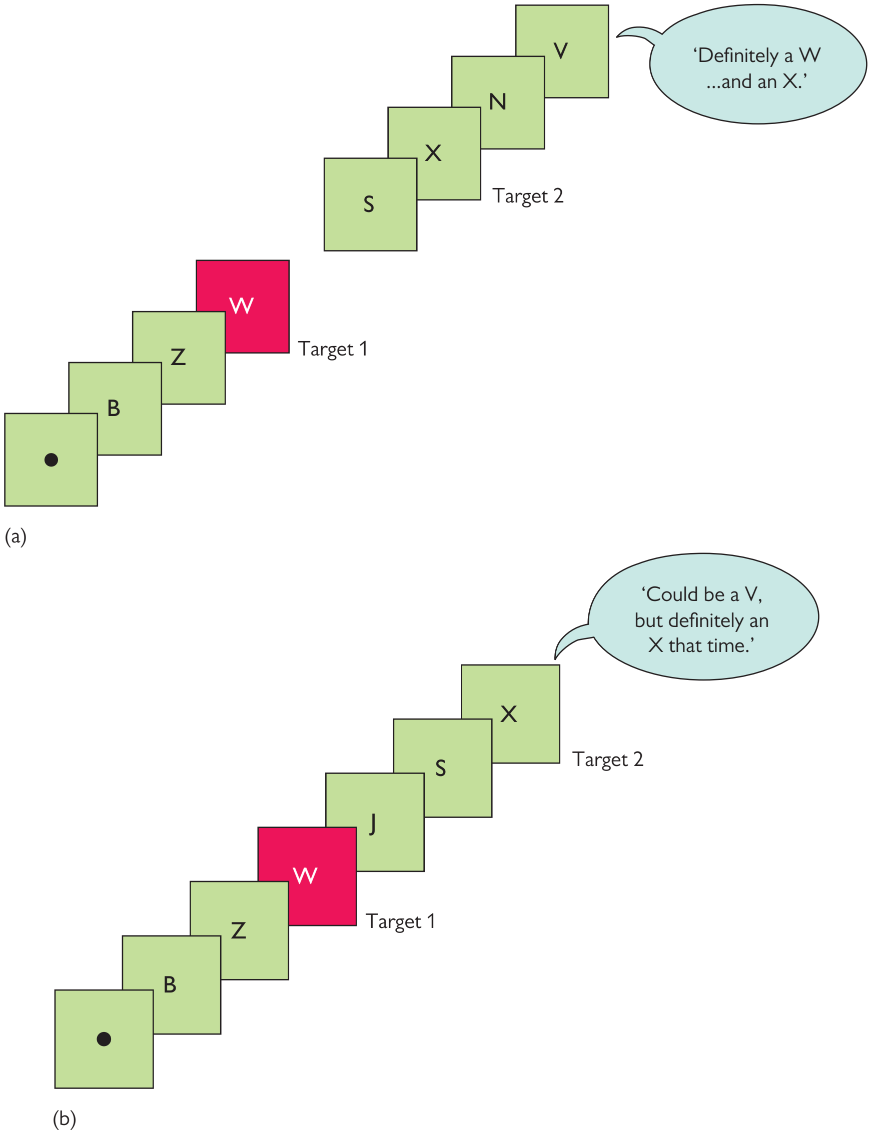

V
N
X
S
'Definitely a W ...and an X.'
Target 2
W
Z
B
●
Target 1
(a)
'Could be a V, but definitely an X that time.'
X
S
J
W
Z
B
●
Target 2
Target 1
(b)

FIGURE 2.4 (a) Target 2 is seen more easily when Target 1 is made easier to see by removing the following item; (b) Target 2 is also seen easily when items following it are omitted.

mask for one target lead to an even larger improvement for another target that was yet to be presented? To return to our earlier analogy, if the shopkeeper is having some trouble in dealing with the first customer, then the second is kept waiting and suffers. That doesn't explain *how* the waiting queue suffers (if it were me I should probably chat to the person behind and forget what I had come for), but that question was also addressed by removing items from the sequence.

Giesbrecht and Di Lollo (1998) removed the items following T2 so that it was the last in the list; again, the AB effect disappeared (see Figure 2.4(b)). So, no matter what was going on with T1, T2 could be seen if it was not itself masked. To explain this result, together with the fact that making T1 easier to see also helps T2, Giesbrecht and Di Lollo developed a two-stage model of visual processing. At Stage 1, a range of information about target characteristics is captured in parallel: identity, size, colour, position, and so on. In the second stage, they proposed, serial processes act upon the information, preparing it for awareness and report. While Stage 2 is engaged, later information cannot be processed, so has to remain at Stage 1. Any kind of disruption to T1, such as masking, makes it harder to process, so information from T2 is kept waiting longer. This has little detrimental impact upon T2 unless it too is masked by a following stimulus (I don't forget what I came to buy if there is no-one else in the queue to chat with). When T2 is kept waiting it can be overwritten by the following stimulus. The overwriting process will be damaging principally to the episodic information; an item cannot be both white and black, for example. However, semantic information may be better able to survive; there is no reason why *shoe* and *rope* should not both become activated. Consequently, even when there is insufficient information for Stage 2 to yield a fully processed target, it may nevertheless reveal its presence through priming or EEG effects. There is an obvious similarity between this account and Coltheart's (1980) suggestion: both propose the need to join semantic and episodic detail.

So often, nice explanations need to be reconsidered in the light of other findings; there is an example of this in the RSVP literature. Shapiro *et al.* (1997) tried putting the participant's own name in the T2 position and, rather like the auditory 'own name effect', their name was detected without difficulty as if there was no attentional blink. In the auditory analogue (Section 1.3) we saw the effect explained by suggesting that unattended material was attenuated but still 'loud' enough to trigger something as well-primed as one's name. However, if you have followed the story so far, you will no longer think of attention as working like a volume control. Rather, it seems to be the 'glue' that unites the different elements of a stimulus and so enables us to be aware of it. Does our own name not require 'gluing'? We will try to answer this question when we come to look at yet another of Broadbent's studies, in Section 3.3.

2.4 Masking and attention

Before I summarize the material in this section and we move on to consider attentional processes with clearly-seen displays, it would be appropriate to consider the relevance of the masking studies to the issue of attention. We began the whole subject by enquiring about the fate of material that was, in principle, available for processing, but happened not to be at the focus of attention. Somehow we have moved into a different enquiry, concerning the fate of material that a participant was trying to attend to, but did not have time to process. This seemed a natural progression as the chapter unfolded, but are the two issues really related? Merikle and Joordens (1997) addressed this very question; they characterized it as a distinction between perception without awareness (such as in masking studies) and perception without attention (as with dichotic listening). They carried out a number of studies in which processing was rendered difficult either by masking, or by giving the participants two tasks so that they could not focus on the target. They concluded that the results were entirely comparable, and that the same underlying processes are at work in both kinds of study.

However, over the space of a decade a different kind of evidence has become available: that from brain scanning. Scanning techniques can reveal the impact of a stimulus 'downstream' as the information moves away from the primary sensory areas of the brain to receive further processing in more distant regions. This kind of information makes it very clear that we should distinguish between those stimuli that we don't see as a result of masking and those that we miss through inattention. In the masking situation the 'invisible' item is generally at the focus of attention, and the only reason that we fail to see it (consciously) is that the information is available for only an exceedingly short time. This can barely have an impact upon the sensory system and, sure enough, brain scanning shows the influence of the stimulus 'fizzling out' as the information moves on for further analysis. In contrast, clear stimuli that do not receive attention do at least succeed in activating many regions of the brain. Kouider and Dehaene (2007) see this as an important distinction, and we shall have to return to this theme later.

SUMMARY OF SECTION 2

The results of the visual attention experiments we have considered can be interpreted as follows.

- Attention can be directed selectively towards different areas of the visual field, much as it can be directed to different sounds.
- The inability to report much detail from brief, masked visual displays appears from brain scanning evidence to be linked to the need to assemble the various information components for an item to be visible.
- Where visual information is available it is captured in parallel, but assembly is a serial process.
- Episodic detail (e.g. colour, position) is vulnerable to the passage of time or to 'overwriting' by a mask.
- Semantic information (i.e. identity/meaning) is relatively enduring but does not reach conscious awareness unless bound to the episodic information.
- Attention, in this context, is the process of binding the information about an item's identity to its particular episodic characteristics.
- 'Unbound' semantic activation can be detected by priming and electrophysiological techniques.

3 INTEGRATING INFORMATION IN CLEARLY-SEEN DISPLAYS

The binding of features emerges as being a very significant process when displays are brief, because there is so little time in which to unite them. With normal viewing, such as when you examine the letters and words on this page, it is not obvious to introspection that binding is taking place. However, if, as explained above, it is a necessary precursor to conscious awareness, the process must also occur when we examine long-lived visual displays. Researchers have attempted to demonstrate that the binding process does indeed take place.

3.1 Serial and parallel search

Examine the three sections of Figure 2.5 and in each case try to get a feel for how long it takes you to find the 'odd one out'.

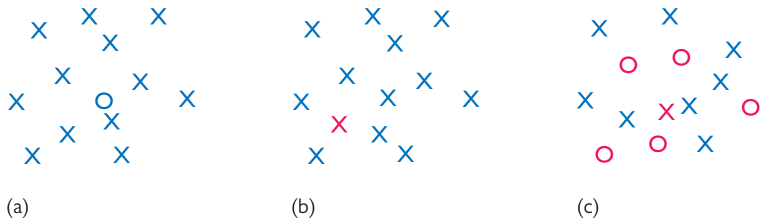


FIGURE 2.5 Typical stimuli used in Treisman's experiments. Find the odd item in each of the groups (a), (b), and (c).

You probably felt that the odd items in Figures 2.5(a) and 2.5(b) simply 'popped out' and were immediately obvious, whereas the red X in Figure 2.5(c) took you slightly longer to find. These kinds of effect have been explored formally by Treisman (e.g. Treisman and Gelade, 1980). The odd item is referred to as the target and the others as the distractors. Treisman showed her participants a series of displays of this nature and measured how long it took them to decide whether or not a display contained a target. She was particularly interested in the effect of varying the number of distractors surrounding the targets. It was found that for displays similar to Figures 2.5(a) and 2.5(b) it made no difference to decision times whether there were few or many distractors. In contrast, with the 2.5(c) type of display, participants took longer to decide when there were more distractor items; each additional distractor added approximately 60 ms to the decision time.

How is that pattern of results to be explained? Treisman pointed out that the first two displays have target items that differ from the rest on only one dimension; the target is either a round letter (O), among 'crossed-line' letters (X), or a red letter among blue letters. The 2.5(c) display type is different; to identify the target it is necessary to consider two dimensions. It has to be an X (but there are others, so on its own being an X does not define the target) and it has to be red (but again, there are other red letters). Only when X and red are combined does it become clear that this is an 'odd one out'. All these features (various colours and shapes) are quite simple and are derived in the early stages of visual processing, but importantly different types of analysis (e.g. of shape or colour) take place in different parts of the brain. To see whether there is just 'redness' or just 'roundness' in a display is easy, so easy in fact that the whole display seems to be taken in at a glance, no matter how many items there are. In other words, all the different items are processed at the same time, in parallel. The situation is very different when shape and colour have to be combined because they are determined in different brain areas; somehow the two types of information have to be brought together. You will recall from Section 2 that attention appears necessary to unite episodic and semantic information. Treisman proposed that it is also required to link simple features. Each item in the display has to receive attention just long enough for its two features (shape and colour) to be combined, and this has to be done one item at a time until the target is found. In other words, the processing is serial, so takes longer when there are more items to process.

It has been known for some time that the **parietal** region of the brain (part of the cortex that sits like a saddle across the top of the brain) is one of the areas involved in attention. A fuller account of the problems that result from damage to this area will be given in Section 5.1; at this point it is relevant to mention that Treisman (1998) reports investigations with a patient who had suffered strokes in that region. This patient was shown simple displays containing just two letters from a set of three (T, X, and O); they were printed in different colours, again from a choice of three (red, blue, or yellow). He was asked to describe the first letter he noticed from the two presented. On a particular occasion he might be shown a blue T and a red O. Although he often made mistakes, he would rarely respond 'Yellow X' to that display; that is, he did not claim to see features that were not there at all, so he was not simply guessing. What he did say quite often would be something like 'Blue O'. He had correctly identified features that were present but was unable to join them appropriately. The implication of this is that both the detection and the integration of features are necessary steps in normal perception, and that integration requires attention.

3.2 Non-target effects

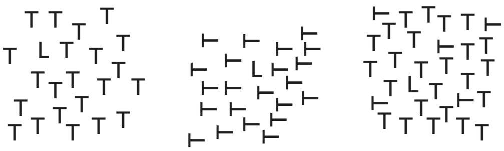
Treisman's **feature integration theory** has been very influential but it does not appear to explain all experimental observations, and there have been alternative accounts of the feature-binding process. Duncan and Humphreys (1989) reported effects that do not fit too well within the basic Treisman account. They required participants to search for the letter 'L' (the target) within a number of 'Ts' (the non-targets). You may get a feel for the relative difficulty of different versions of their task by examining Figure 2.6.

The task can be conceptualized as looking for two lines that meet at a corner (the L), rather than forming a T-junction. It should not make much difference whether the T-junctions are vertical or horizontal (as in Figure 2.6(a) and 2.6(b)) and, indeed, the search times for these two sorts of display are similar. However, when the Ts are mixed, as in Figure 2.6(c), it takes longer to find the target. This finding would not have been predicted by a simple feature integration theory. Duncan and Humphreys (1989) argued that part of finding the target actually involves rejecting the non-targets and that this is a harder task when they come in a greater variety.

This explanation does not rule out the idea that features need to be integrated to achieve recognition, but it does suggest that non-targets as well as targets need to be recognized. The following section also describes evidence that non-targets are recognized, but in this case the recognition appears to take place in parallel. That really does deserve closer examination, because, if true, it would imply that all the previous evidence for *serial* feature combination may have been misinterpreted.

3.3 The 'flanker' effect

A potential problem for the feature integration theory is the fact that the time taken to understand the meaning of a printed word can be influenced by other, nearby words. In itself this is not surprising, because it is well known that one word can prime (i.e. speed decisions to) another related word; the example *nurse – doctor* was given in Section 1.3. However, Shaffer and LaBerge (1979) found priming effects even when they presented words in a way that might have been expected to eliminate priming. For their experiment a word was presented on a screen, and as quickly as possible a participant had to decide to what category it belonged; for example, an animal or a vegetable. The participant was required to press one button for animal names and another for vegetables. This sounds straightforward, but the target word was not presented in isolation; above and below it another word was also printed, making a column of three words. The target, about which a decision was to be made, was always in the centre. The words repeated above and below the target were termed the 'flankers'. Before the three words were displayed, markers in the field of view showed exactly where the target would appear. Figure 2.7 shows examples of possible displays.



(a) (b) (c)

FIGURE 2.6 Examples of the kinds of stimuli used by Duncan and Humphreys (1989). Find the letter L in each of the groups (a), (b), and (c).

cat	pea
dog	dog
cat	pea
(a)	(b)

FIGURE 2.7 The flanker effect. It takes longer to decide 'dog' is an animal when it is surrounded by words of another category, as in (b).

You will probably not be surprised to learn that people make category judgements more quickly for examples such as that shown in Figure 2.7(a) than for the 2.7(b) type of stimulus. Presumably, while the target information is being processed, details about the flankers are also being analysed, in parallel. When they turn out to be from the category associated with pressing the other button they slow the response. This slowing is very much like the impact of the conflicting colour names in the Stroop effect (see Box 2.2). However, recall that Treisman's theory suggests that focused, *serial* attention is required to join features together. A printed word has many features and it would be thought that they require joining before the word can be recognized; it should not be possible to process the three words simultaneously. A participant focusing on the target could not (according to the theory) also be processing the flankers.

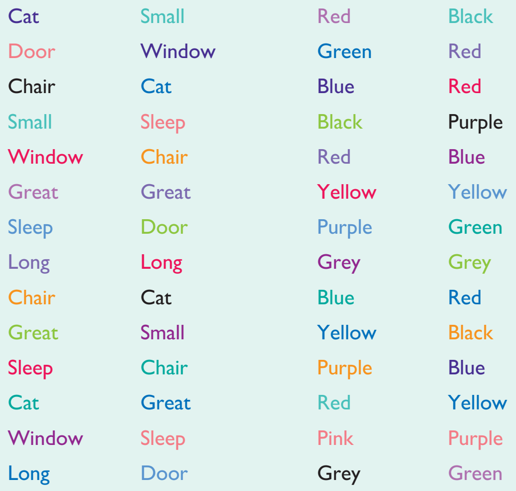
Broadbent addressed this problem (Broadbent and Gathercole, 1990) and produced an explanation to 'save' the feature integration theory. He suggested that the central target word primed the flankers so effectively that they could be detected with the minimum of attention. Taking the items in Figure 2.7 as an example, if this explanation were true it would have to be argued that 'dog' primes 'cat', which being another animal leads to faster decision times. 'Dog' cannot prime 'pea', as they are unrelated, so there is nothing to make the decision any quicker. In other words, it is not that 'pea' makes responses to 'dog' harder; rather, 'cat' makes them easier. Broadbent and Gathercole tested this explanation with an ingenious modification to the usual way of presenting targets and flankers. Instead of displaying all three words simultaneously, the target appeared first, to be joined by the flankers 40 ms later. The sequence is represented in Figure 2.9.

The reasoning behind this change was as follows. If Broadbent and Gathercole were correct that the flankers were analysed only because of priming from the target word, then giving the target a 'head start' should

BOX 2.2 RESEARCH STUDY The Stroop effect

Stroop (1935) reported a number of situations in which the processing of one source of information was interfered with by the presence of another. The best known example uses a list of colour names printed in non-matching coloured inks (see Figure 2.8).

A variant is the 'emotional Stroop task', which can be used in therapeutic diagnoses. For example, severe depression produces cognitive impairment and, in the elderly, it is difficult to distinguish this from the effects of the onset of dementia. Dudley *et al.* (2002) used colours to print a list of words, some of which were associated with negative emotions (e.g. the word sadness). Depressed people have an attentional bias towards such depression-related material. Patients were required to name the ink colours for each word as quickly as possible. Both depressed patients and those in the early stages of Alzheimer's disease were slower than a control group, but only the patients with depression were extra slow in responding to negative words. The technique permits an appropriate diagnosis.



Cat
Door
Chair
Small
Window
Great
Sleep
Long
Chair
Great
Sleep
Cat
Window
Long

Small
Window
Cat
Sleep
Chair
Great
Door
Long
Cat
Small
Chair
Great
Sleep
Door

Red
Green
Blue
Black
Red
Yellow
Purple
Grey
Blue
Yellow
Purple
Red
Pink
Grey

Black
Red
Red
Purple
Blue
Yellow
Green
Grey
Red
Black
Blue
Yellow
Purple
Green

FIGURE 2.8 The Stroop effect. The task is, *as quickly as possible*, to name the colour in which each word is printed; do *not* read the words. Try the first two columns to start with – the task is not too difficult. Then try the next two columns. The conflicting colour words are very disruptive.

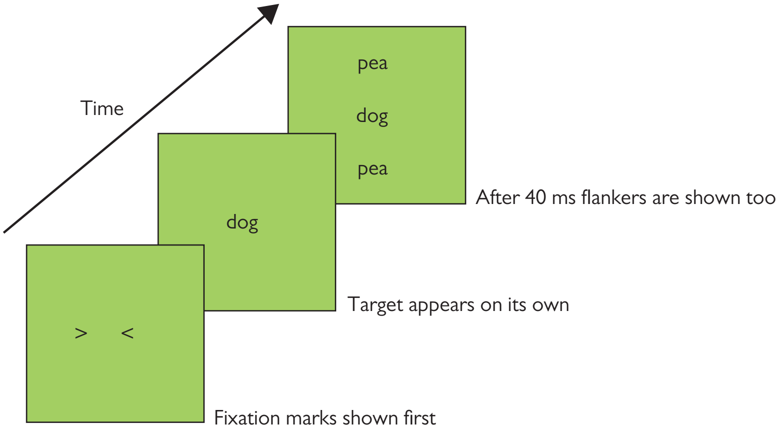

Time
pea
dog
pea
After 40 ms flankers are shown too
dog
Target appears on its own
> <
Fixation marks shown first

FIGURE 2.9 The Broadbent and Gathercole (1990) modification: the flankers are delayed for 40 ms.

enable it to prime even more effectively; the flanker effect would be even stronger. On the other hand, if interference from the flankers was merely an example of processing not being as 'serial' as Treisman supposed, then making flankers arrive late, when target processing had already started, should *reduce* the flankers' impact. Well, the delayed flankers produced a strong effect (i.e. faster responses with same-category flankers), suggesting that the priming idea was correct. However, there is another interpretation of the Broadbent and Gathercole results. It has been well established that an item suddenly appearing in the visual field will capture attention (e.g. Gellatly *et al.*, 1999). By making the flankers appear later, Broadbent and Gathercole may have ensured that they would attract attention away from the target. This could explain why the flankers showed a particularly strong effect with this style of presentation. Although the Broadbent and Gathercole idea of staggering the display times of the stimuli was ingenious, to investigate parallel processing convincingly requires all the different stimuli to be presented at the same time.

Gronau *et al.* (2003, 2009) have used a variant of the Stroop task to investigate the extent to which simultaneously presented, unattended material can influence response times. They used coloured words and rectangles, all presented at the same time, in somewhat the same way as with the flanker task. The target was a coloured rectangle, presented on a screen, and participants had to indicate, as quickly as possible, what the colour was. These researchers tested a large range of variations on their basic idea, but the underlying concept and findings were as follows. Inside or outside the rectangle a word (two words if on the outside, i.e. like the flanker stimuli) was also presented. If the words were colour names, different from the rectangle colour, then the responses were slower, just as with the conventional Stroop test. As well as colour words another type of word was sometimes used – people's names. You will recall that the emotional Stroop test (Box 2.2) detects slowing caused by emotionally salient words. One's own name is just such a word; if you try to name the colour used to print your name you take longer than if it is someone else's name. So, Gronau *et al.* used names, either inside or outside the rectangle, and sometimes the name was the participant's own. Just as with the conflicting colour words, the personal name caused slowing, *but only when it was presented inside the rectangle*.

What are we to make of that? To recap, colour name words caused typical Stroop interference, whether inside or outside the target rectangle, but personal names interfered only if inside. The difference between these classes of word is that colour names were relevant to the task, which was itself about naming colours; it was not about people's names. The results tell us something about the narrowness of the focus of attention. The task was to name the colour of the rectangle, so attention was focused upon that. Inevitably, anything within the rectangle would fall under the 'spotlight of attention' too, and that is why both the colour words and the names caused slowing when they were presented within the rectangle. Personal names did not slow responses when they were outside the rectangle, so that region was obviously outside the area of attention. However, the colour words *did* interfere, even when in the unattended area, so they must have been processed without requiring any attention. It would seem that items that are relevant to the task in hand (colour naming here) can be processed without attention, whereas irrelevant material, such as personal names, cannot be processed unless it receives attention.

The above finding suggests that Broadbent and Gathercole were wrong to think that priming was necessary to explain the processing of the flankers; they were processed because they were relevant to the task, i.e. deciding upon category membership. This does leave us with two questions, and we shall have to return to these. First, exactly how does the relevant material become processed if attention is normally required to unite ('glue together') the stimulus components? The second question raises a different sort of issue: why does one's name not get processed when unattended, if the 'own name' results (Section 1.3) imply that it is sometimes processed in exactly those circumstances?

SUMMARY OF SECTION 3

When consciously perceiving complex material, such as when looking for a particular letter of a particular colour:

- Perception requires attention.
- The attention has to be focused upon one item at a time, thus . . .
- processing is serial.
- Some parallel processing may take place, but . . .
- it is detected indirectly, such as by the influence of one word upon another.

4 ATTENTION AND DISTRACTION

The above account of attention spreading beyond the intended target reminds us that, while it may be advantageous from a survival point of view to have attention captured by peripheral events, these events are actually distractions from the current object of attention. Those who have to work in open-plan offices, or try to study while others watch TV, will know how distracting extraneous material can be. Some try to escape by wearing headphones, hoping that music will be less distracting, but does that work? Are some distractors worse than others? These kinds of question have been addressed by research and the answers throw further light upon the nature of attention.

4.1 The effects of irrelevant speech

Imagine watching a computer screen on which a series of digits is flashed, at a nice easy rate of one per second. After six items you have to report what the digits had been, in the order presented (this is called serial recall – see also Chapters 9 and 19). Not a very difficult task, you might think, but what if someone were talking nearby? It turns out that, even when participants are instructed to ignore the speech completely, their recall performance drops by at least 30 per cent (Jones, 1999).

In the context of dichotic listening (Section 1.2) it was shown that ignored auditory material may nevertheless be processed, and hence its meaning influences perception of attended material. However, meaning appears to have no special impact when speech interferes with memory for visually presented material. Thus, hearing numbers spoken while trying to remember digits is no more damaging than listening to other irrelevant speech items (Buchner *et al.*, 1996). In fact, even a foreign language or English played backwards are no less disruptive than intelligible speech (Jones *et al.*, 1990). On the other hand, simple white noise (a constant hissing like a mistuned radio) is almost as benign as silence. Interference presumably results from speech because, unlike white noise, it is not constant; it is broken into different sounds.

The importance of 'difference' in the speech can be shown by presenting lists of either rhyming or non-rhyming words. It turns out that a sequence such as 'cat, hat, sat, bat . . .' is less disruptive than a sequence such as 'cat, dog, hit, bus . . .' (Jones and Macken, 1995). Jones (1999) proposes that, whether listening to speech, music, or many other types of sound, the process requires the string of sounds to be organized into perceptual 'objects'. To recognize an auditory object, such as a word or melody, requires that the segments of the stream of sounds are identified, and it is also necessary to keep track of the order of the segments. This ordering process, which occurs automatically, interferes with attempts to remember the order of visually presented items. When the sounds contain simple repetitions (as with the rhyming 'at' sound) the ordering becomes simpler, so the memory task is less disrupted. This was demonstrated in a surprising but convincing way by Jones *et al.* (1999). Their participants attempted to remember visually presented lists while listening through headphones to a repeating sequence of three syllables, such as the letter names 'k . . . l . . . m . . . k . . . l . . . m'. These were disruptive, since the three letters have quite different sounds. The experimenters then changed the way in which the speech was delivered. The 'l' was played through both headphones, so sounded in the middle (see Section 1.2, Box 2.1) but the 'k' was played only to the left ear and the 'm' was heard in the right. This manipulation resulted in the perception of three 'streams' of speech, one on the left saying 'kay, kay, kay . . .', one in the middle repeating 'ell', and the last on the right saying 'em'. The significant point is that instead of hearing a continually changing sequence, the new way of playing *exactly the same sounds* resulted in them sounding like three separate sequences, each of which never changed. Remarkably, the result was that they were no longer as disruptive to the visual recall task.

This section has taken the concept of attention into a new area. Previously we have seen it as a means of separating information or of directing the assembly of different aspects of the attended item. In most of the earlier examples it has appeared that a great deal of processing can take place in parallel, although the results may not all reach conscious awareness. The impact of irrelevant speech shows that parallel processing is not always possible. It seems to break down in this case because demands are made on the same process – the process that places items in a sequence. Here it would seem that we have a situation where there really is a 'bottleneck' of the sort envisaged in early theories of attention (see Sections 1.2 and 1.3).

What of trying to study with music? Undoubtedly 'silence is golden', but if music is to be played then my suggestion is that it should perhaps be something that changes very slowly, such as the pieces produced by some of the minimalist composers.

4.2 Attending across modalities

The preceding section raised the issue of attention operating (and to some extent failing) across two sensory modalities. By focusing on distraction we ignored the fact that sight and sound (and other senses) often convey mutually supporting information. A classic example is lip-reading. Although few of us would claim any lip-reading skills, it turns out that, particularly in noisy surroundings, we supplement our hearing considerably by watching lip movements. You can observe this at a nightclub (the music always seems to be at an ear-damaging intensity) where people trying to hold a conversation watch each other's faces very closely. If attention is concerned with uniting elements of stimuli from within one sense, then we might expect it to be involved in cross-modal (i.e. across senses) feature binding too. In this section we will look briefly at one such process.

A striking example of the impact of visual lip movements upon auditory perception is found in the **ventriloquism effect**. This is most commonly encountered at the cinema, where the loudspeakers are situated to the side of the screen. Nevertheless, the actor's voice appears to emanate from the face on the screen, rather than from off to the side. Driver (1996) demonstrated just how powerful this effect could be. He presented participants with an auditory task that was rather like shadowing in dichotic listening (Section 1.3) – only much harder! The two messages, one of which was to be shadowed, did not go one to each ear; they both came from the same loudspeaker and were spoken in the same voice. To give a clue as to which was to be shadowed a TV monitor was placed just above the loudspeaker, showing the face of the person reading the message to be shadowed. By lip-reading, participants could cope to some extent with this difficult task. Driver then moved the monitor to the side, away from the loudspeaker. This had the effect of making the appropriate message seem to be coming from the lips. Since the other message did not get 'moved' in this way, the two now *felt* spatially separate and, although in reality the sounds had not changed, the shadowing actually became easier!

These kinds of effects have further implications at a practical level. The use of mobile telephones while driving a car has been identified as dangerous, and the danger is not limited to the case where the driver tries to hold the phone in one hand, change gear with the other, and steer with the knees! If a hands-free headset is used of the type that delivers sound via an earpiece to just one ear, the caller's voice sounds as if it is coming from one side. Attending to this signal has the effect of pulling visual attention towards the lateral message, reducing the driver's responsiveness to events ahead (Spence, 2002).

SUMMARY OF SECTION 4

We have seen that attentive processes will 'work hard' to unite information into a coherent whole.

- Even spatially separate visual and auditory stimuli can be joined if they appear to be synchronous (the ventriloquism effect).
- When stimuli are not synchronous the system attempts to order the segments of the stimuli independently, resulting in distraction and lost information.
- It is a 'bottleneck' in the ordering process that results in one stream of information interfering with the processing of another.

5 THE NEUROLOGY OF ATTENTION

Modern techniques for revealing where and when different parts of the brain become active have recently provided a window on the processes of attention. For example, one of these brain-scanning techniques, functional magnetic resonance imaging (fMRI), has been used to show the behaviour of an area of the brain that responds to speech. It turns out also to become activated in a person viewing lips making speech movements *in the absence of sound*. For this to happen there must be connections between relevant parts of the visual and auditory areas.

5.1 The effects of brain damage

Before the advent of 'brain mapping', such as by fMRI, it was nevertheless possible to discover something of the part played by different regions of the brain by observing the problems resulting from brain damage (such as following a stroke). One such area was mentioned in Section 3.1 – the parietal lobe. Damage to a single lobe (there is one on either side) leads to what is called **sensory neglect**, or sometimes simply neglect. Generally the effect is worse when the damage is to the right hemisphere; this side seems to be particularly involved in visuo-spatial awareness. A patient with this deficit is likely to ignore the doctor completely if s/he stands on the neglected side (the side opposite to the site of the damage). When eating, the patient will probably leave any food that is on the 'wrong' side of the plate, and if asked to draw a flower will put petals on only one side. The problem is not simply blindness to all that lies on the neglected side. A patient asked to draw a whole vase of flowers may draw only those hanging over the 'preserved' side, but with each individual flower itself only half complete. It appears sometimes to be half the *object* that is neglected, rather than half the field of view. Figure 2.10 shows a typical attempt by a patient with visual neglect to draw a clock face.

That neglect may be associated with the object rather than the scene was demonstrated formally by Driver and Halligan (1991). They showed patients pairs of pictures that looked rather like silhouettes of chess pieces. Patients had to say whether the two pictures were the same or different. Where there *were* differences, they comprised an addition to one side, near the top of the figure (as if the chess queen had something attached to one ear!). When the addition was on the neglected side patients were unable to detect the difference. Suppose the 'problem' side was the left. The question is whether the patient has difficulty with processing information to the left of the page or to the left of the object. Driver and Halligan tested this by tilting the pictures to the right (see Figure 2.11), so that the one-sided feature, although still on the left of the figure, was now in the right half of the page. Still the patients experienced difficulty: neglect was object-related.

In addition to being linked to objects, there appears to be an action link to attention. We will return to this theme in Section 6, but here it is relevant in the context of neglect. Humphreys and colleagues (Humphreys *et al.*, 2004; Riddoch *et al.*, 2003) have explored the responses of patients to items that 'go together', such as a wine bottle and corkscrew. The normal tendency for patients presented with two pictures, on either side of

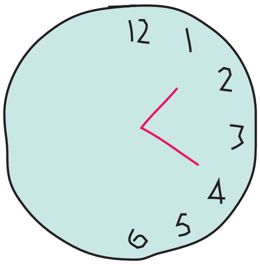


FIGURE 2.10 The typical appearance of a clock face, as drawn by a patient with visual neglect.

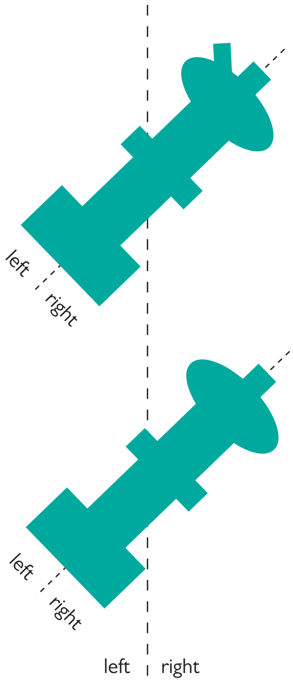


FIGURE 2.11 Same or different? The feature that distinguishes the two figures is to the left of the object, but on the right of the page.

the mid-line, is for them to report just one, from the spared field of view. However, they did much better when the pictures would, in real life, be linked by an action, such as a bottle and corkscrew, which would of course be screwed into the top of the bottle. This advantage was obtained when the screw was actually pointed at the bottle top, as if ready for use, but lost if it was positioned to point at the base of the bottle. It would seem that seeing objects in a position that implied their normal use somehow served to combine them, so that the part of the combination that was in the neglected field of view could be 'rescued'.

We have been describing attention as a mechanism for assembling the sub-components of items in a scene, so it is not difficult to conceptualize a fault leading to some components being omitted. This account sees attention as an essential element of the perceptual process, helping to organize incoming information. However, neglect is not limited to objects that are physically present. Bisiach and Luzzatti (1978) asked their patient to imagine standing in the cathedral square of the Italian city where he grew up. He was to imagine looking towards the cathedral and to describe all that was in the square. He did this very well, except that he failed to mention any of the buildings down the left-hand side of the square (his brain injury was on the right). He was then asked to imagine standing on the cathedral steps, looking back towards his previous viewpoint. Again, he only reported details from the right. However, with the change of view, this meant that he was now describing previously ignored buildings! Clearly his memory was intact, but in some way not entirely accessible. Equally clearly, attentive processes are involved in the assembly of remembered material as well as of physically present stimuli.

An even more extreme form of neglect is encountered in a condition known as Balint's syndrome. It occurs when a patient is unfortunate enough to suffer damage to both parietal lobes, which results in it being extremely difficult to shift attention from one object to another. Thus, when trying to light a cigarette, the patient may find that their attention has been 'captured' by the flame, to the extent that they can no longer see the cigarette. One patient complained, 'When I see your spectacles I cannot see your face.' This is reminiscent of the experience of pilots using a head-up display (HUD) (see Section 2) where focusing on flight information displayed in the HUD makes the outside scene feel less 'visible'. Surprising as it may sound, it seems necessary to deduce from these effects that we *all* experience the world as a series of objects. However, unless our attentive process has been damaged, we can shift the attention so rapidly from one object to another that we perceive them all as being present simultaneously. Exactly what constitutes an object depends upon the situation; Balint patients are revealing here because they see only one object at a time. Baylis *et al.* (1994) described a patient who could not report the letters making up an isolated word. Viewed in this way, each letter was a small object and it was not possible to switch attention from one to the next. However, the patient could read the whole word, since for this purpose it was a single object.

Early visual processing takes place in two major pathways in the brain known as the ventral and dorsal streams (these are described in Section 6 of Chapter 3); the parietal region is part of the dorsal pathway. Damage to the ventral stream results in different kinds of integration problems; patients are aware of all aspects of a scene but to the patient they remain segmented into small elements. For example, an individual shown a photograph of a paintbrush described seeing a wooden stick and a black object (the bristles), which he could not recognize. Humphreys (2001) suggests that the varieties of different problems are evidence that the binding together of different features takes place in several different stages and brain locations. This is underlined by recent work of Humphreys and colleagues (Humphreys *et al.*, 2010). They built upon the finding that picturing objects as if ready for use saved them from being missed by neglect patients. The extent of this saving has now been shown to be even more marked when the objects are correctly located for the patient's personal way of using them. For example, a left-handed patient would perform best if a teapot were pictured on the left with the cup to the right, because they would have held a real teapot in their left hand. The authors conclude that, in addition to the visual-based grouping achieved by placing objects appropriately with respect to each other, there is another grouping that can be thought of as motor-based (i.e. using movement), which comes into play when objects are placed appropriately for the viewer.

5.2 Event-related potentials

When a sense organ (eye, ear, etc.) receives a stimulus, the event eventually causes neurons to 'fire' (i.e. produce electrical discharges) in the receiving area of the brain. The information is sent on from these first sites to other brain areas. With appropriate apparatus and techniques it is possible to record the electrical signals using electrodes attached to the scalp. The electrical potentials recorded are called **event-related potentials (ERPs)** because they dependably follow the triggering sensory event. In fact a whole series of electrical changes are detected, first from the receiving brain areas and then later from subsequent sites. The timing of the ERPs gives a clue as to where in this sequence they are being generated.

Woldorff *et al.* (1993) examined ERPs evoked by sounds. These included signals occurring as soon as 10 ms after the auditory event. To generate a response so quickly, these ERPs must have originated in the brain stem, in the first 'relay' between ear and auditory cortex. The earliest stages of registration at the auditory cortex were detected after about 20–50 ms. It was of particular interest that, whereas the 10 ms signal was not affected by attention, the magnitude of the electrical activity in the cortex was smaller when the sounds were played to an unattended ear. This shows that, at a very early stage of cortical analysis, attending away from a stimulus actually reduces the intensity of the signal in the brain. The result lends a good deal of support to the theory that attention is exercised by controlling a filter early in the processing sequence (see Section 1.3). Note, however, that the unattended signal is only attenuated, not eliminated.

SUMMARY OF SECTION 5

Many familiar themes have re-emerged in this section, together with the recognition that attention is involved in the assembly of remembered material as well as of current perceptions.

- Attention is associated with the generation of perceptual objects.
- When objects are used together they can be treated as if they are one item.
- In addition to being an essential part of external stimulus processing, attention influences remembered experiences.
- ERP data show that cortical signals derived from unattended external stimuli are attenuated.

6 DRAWING IT ALL TOGETHER

Perhaps that's what attention does – draws together the components of a stimulus, but before jumping to conclusions we must try to draw together the disparate elements of this long account. At this stage it will be helpful to consider briefly the issue from new perspectives – from viewpoints that are not always associated with attention. I will allow myself to be a little speculative too!

We seem to have come a long way and covered a great deal of ground since I approached this subject by explaining that a mechanism must exist to help us focus on one sound out of many. That clearly is one function of attention, but attention seems to have other functions too. The results of visual search experiments show that attention is a vital factor in joining together the features that make up an object, and the experiences of brain-damaged patients suggest that this feature-assembly role ensures that our conscious perceptions are generally of objects rather than of their constituent parts. Cross-modal research has demonstrated that the gathering together of related information from different senses is also controlled by attention.

Attention has a role to play in dealing with competition. The early researchers believed that attention was vital because the brain would be able to deal with only one signal at a time; a 'winning' signal had to be picked from among the competitors. Although we have shown that a good deal of analysis can actually take place in parallel, there are also results that suggest that more complex analysis is largely serial, thus requiring a mechanism to select from the competing stimuli. Often, the parallel processes have to be demonstrated rather obliquely, since their results do not become consciously available. Thus attention has to do with what reaches conscious awareness. Why should this be so? Why should we not be equally aware of several items simultaneously?

Allport (1987) offered an answer that suggests yet another role for attention: as hinted at in Section 5.1, it is to direct actions. Although we might, in principle, be able to perceive many things at once, there are situations where it would be counterproductive to attempt to *do* more than one thing. Allport gave fruit-gathering as an example. When we look at a bush of berries we need to focus attention upon one at a time, since that is how they have to be picked. If animals had not evolved this ability to select, if all the food items remained equally salient, they would starve as they hovered over them all, unable to move toward any one! From this perspective, attention is the process that saves us from trying to carry out incompatible actions simultaneously. However, everyday experience reminds us that the issue of consciousness remains relevant. For example, novice drivers experience considerable difficulty in trying simultaneously to perform all the actions needed to control a vehicle; in Allport's view they are trying to 'attend-for-action' to more than one thing at a time. However, this could be restated as an attempt to be *conscious* of more than one thing at a time. Once the driver has become more skilful, the difficulty of combining actions disappears, but so too does the driver's conscious awareness of performing them: they have become automatic.

BOX 2.3 RESEARCH STUDY Hypnosis and the hemispheres

Naish (2010) used a temporal order judgement (TOJ) task to explore hemispheric effects in hypnosis. TOJ is a task in which two events occur and a participant has to say which came first. If the events are short and the gap between them is brief it becomes impossible to judge which was first, so the gap is adjusted to find the amount of time that is required to be just able to do the task. Naish used flashes of light from two small light-emitting diodes (LEDs) as the two events. The LEDs were positioned on either side of the face, near the outer corners of the eyes. In these positions the light stimuli would each be processed in one hemisphere of the brain; because of the contralateral arrangement of the brain, the left light would be 'seen' in the right hemisphere and vice versa. When the TOJ is used in this situation the gaps needed to detect the order can differ, depending whether the order is left-right or right-left; this is assumed to be because one hemisphere processes faster than the other. Naish tested people who were highly responsive to hypnosis and also those who were non-responsive; testing was carried out both without hypnosis and following a hypnotic induction. Results indicated that people who did not respond to hypnosis had a slight right hemisphere advantage (left-first gap smaller) and this did not change significantly when an attempt was made to hypnotize them. In contrast, hypnotizable participants showed a marked left hemisphere advantage outside hypnosis, which changed to a strong right advantage when they were hypnotized. It is suggested that this cross-over facilitates the processes that generate hallucinations.

One might well ask how the term 'attention' has come to be applied to so many roles and processes; it might have been better to use different labels to distinguish between them. To use one word with so many aspects certainly makes a unitary definition very difficult to formulate. I suspect that the single term has stuck because ultimately all these facets of attention do lead to one result: conscious awareness.

6.1 Attention as the generator of consciousness

Consciousness as a topic is dealt with in Chapter 18, but it is appropriate here to consider its close association with attention. It was mentioned above, and in several other sections of this chapter, that attention is required to unite the elements of a stimulus, but we have never stopped to consider how the unification takes place. Neural oscillations occur across the brain at a range of different frequencies; you may have heard of alpha (α) waves, which can be seen clearly in EEG recordings when a participant relaxes with eyes closed. α-waves have a frequency of about 10 Hz (i.e. ten cycles per second) and seem to arise principally from the visual region of the brain. Another wave type is referred to as gamma (γ), oscillating at about 40 Hz and found widely across the cortex. There is growing evidence that when different brain regions are contributing to the same unified perception their respective γ-waves become synchronized. This is analogous to an orchestra (the brain) where a subset of the players follows the conductor, while the rest 'do their own thing' quietly in the background. The analogy invites the question: Where is the conductor? The answer appears to be in the prefrontal cortex; it contains neurons with long axons, stretching back to the sensory regions of the brain.

Altered states of consciousness cast some light on the effects of γ-wave synchronization and its absence. People suffering from schizophrenia experience hallucinations and feel that their actions are 'not their own' but controlled from outside. There appears to be a breakdown in normal gamma activity in these patients (Haig *et al.*, 2000). Being hypnotized is another example of experiencing changes to consciousness. There are a number of parallels between schizophrenia and hypnosis, the latter also producing hallucinations and the feeling that actions somehow happen 'by themselves'; De Pascalis (2007) describes loss of gamma synchronization in hypnosis too. Naish (2010) has shown that when susceptible people are hypnotized there is a shift in activity towards the right hemisphere (see Box 2.3). There is also evidence that something similar may happen in schizophrenia (Caligiuri *et al.*, 2005). Might the right hemisphere have some special role in generating hallucinations? It was mentioned in Section 5.1 that this hemisphere is more crucially involved in the processes that lead to spatial awareness; Naish (2010) argues that this hemisphere is more likely to take the lead in generating hallucinations. Since, by definition, a hallucination cannot be supported by incoming sensory data, it follows that the experience is a top-down process, presumably initiated by the prefrontal region. How might this come about?

In Section 2.4 the distinction was made between two kinds of stimulus that fail to result in conscious awareness. One kind is the heavily masked stimulus, which, although at the centre of attention, is unable to initiate sufficient neural activity to reach stages that are associated with consciousness. The other kind of stimulus generates ample neural activity but fails to reach consciousness because it is not attended. In what way does that lack of attention disadvantage the signal? The answer is based on the fact that we now know that, at every level of processing, the brain has connections that lead 'backwards' in addition to those that carry the information forwards to the next stage. Thus, the output of every stage appears to feed back into that same stage of analysis, a process that can speed the analysis, produce more distinct output, and, importantly, produce a localized cycle of activity that continues to reverberate for a short while, like a bell that has been struck. This kind of feedback is termed re-entrant processing, and it explains many of the effects of masking. Kouider and Dehaene (2007) use the results of brain scanning to support the idea that very brief, masked stimuli are too weak to initiate the local reverberation. Presumably it is this reverberation that preserves the second target in RSVP (Section 2.3) just so long as it is not changed to a new pattern by a following stimulus. The 'preservation' of T2 is awaiting attention, and that involves the establishing of long-range feedback and reverberation from the prefrontal cortex. This reverberation is much like that of the earlier stages, except in one crucial detail: the feedback that makes it possible is not automatic, it is flexible and depends upon the deployment of attention. Once these long-range circuits are established the activity initiates a conscious experience (Lamme, 2003; Kouider and Dehaene, 2007).

It now becomes clearer why numerous stimuli can receive extensive processing in parallel, while normally only one is successful in reaching conscious awareness. If the fronto-posterior (i.e. front-to-back) reverberation is synonymous with the γ-band synchronization described earlier, then conscious experience would become meaningless if all analysed signals were synchronized. Remember that the idea of a conductor picking out a subset of players was an analogy for uniting the elements that belonged to one stimulus; if everything were synchronized and united we would experience a very strange stimulus indeed! In a sense, attention *imposes* a bottleneck, rather than saving a bottleneck from being overloaded as the earlier researchers had supposed. The width of the neck is variable and depends in part upon what is considered to be a unitary stimulus. The finding that neglect can be defeated when items are in a 'used together' configuration (Section 5.1) suggests that there may be situations where a rather wide range of elements from different stimuli are brought together.

Presumably the synchronizing feedback is deployed strategically and can shift rapidly. Our expectations speed our recognition of words, so it may be assumed that the feedback loops are being established just ahead of the arriving stimulus. In this way it is easy for conscious awareness to pick up an expected story in the 'wrong' ear during dichotic listening (Section 1.3). Similarly, task demands will influence the feedback circuits, so that relevant words not actually at the focus of spatial attention can nevertheless gain some benefit from the synchronizing enhancement and influence decisions (Section 3.3).

The experiencing of neglect when trying to recall a visual scene (Section 5.1) is a reminder of the close links between the processes that bring a current stimulus into conscious awareness, and those that generate recollections or imagined events. It seems likely that the same long-range feedback loops are involved in the two situations; brain scanning certainly shows that the overall patterns of activity during real and imagined experiences are very similar – but not identical. Residual differences may result from the fact that activity due to imagination must be generated entirely from top-down processes. This was proposed earlier for the production of hallucinations, but the results clearly feel more realistic than when simply imagining. Scanning during hypnotically-generated hallucinations shows that whether these are visual (Kosslyn *et al.*, 2000) or auditory (Szechtman *et al.*, 1998), the patterns are virtually identical to the genuine experience. Naish (2007) has suggested that it is the behaviour of a specific part of the frontal brain, the anterior cingulate cortex (ACC), that determines the level of reality experienced, and that the ACC serves to regulate the degree of influence that the prefrontal cortex can exert.

6.2 Conclusions

So there we have it, attention is the process that not only unites the elements that go to make up a stimulus, but also brings the result to conscious awareness. Indeed, uniting and consciousness appear to be one and the same.

Can that be right? A good case can be made for suggesting that consciousness evolved to enable us to 'replay' memories and to imagine, as described above. Baumeister and Masicampo (2010) have taken this a little further and suggest that it is because we were evolving to be a complex social animal that we needed this facility to observe and revisit our own behaviour. The implication is that less complex creatures do not have consciousness as we do (something many people would have assumed anyway). Yet it seems very likely that all but the very simplest must have a mechanism for uniting the components of a stimulus. This implies that there may be at least some animals that can unite the elements of stimuli, yet not have conscious awareness of the result. Perhaps the suggestion made above, that uniting and consciousness are effectively two sides of the same coin, is not true after all. Clearly there is still room for more research in this tantalizing field!

I promised at the start of this chapter that attention was a broad and intriguing topic. I am sure you will agree that it was broad – and we haven't covered half of it, nor reached an entirely satisfactory definition – but I hope you are now intrigued too. It is generally accepted that readers cannot continue to devote attention to text that goes on too long, so I trust that I have stimulated, rather than sated, your attention!

FURTHER READING

Styles, E.A. (1997) *The Psychology of Attention*, Hove, Psychology Press. A very readable textbook, which covers and extends the topics introduced in this chapter.

Pashler, H. (ed.) (1998) *Attention*, Hove, Psychology Press. An edited book with contributors from North America and the UK. Topics are dealt with in rather more depth than in the Styles book.

Kouider, S. and Dehaene, S. (2007) 'Levels of processing during non-conscious perception: a critical review of visual masking', *Philosophical Transactions of the Royal Society, B*, vol.362, pp.857–75. A very readable overview of the development of theories concerned with attending to hard-to-see stimuli.

REFERENCES

Allport, D.A. (1987) 'Selection for action: some behavioural and neurophysiological considerations of attention and action', in Heuer, H. and Sanders, A.F. (eds.) *Perspectives on Perception and Action*, Hillsdale, NJ, Lawrence Erlbaum Associates.

Baumeister, R.F. and Masicampo, E.J. (2010) 'Conscious thought is for facilitating social and cultural interactions: how mental simulations serve the animal-culture interface', *Psychological Review*, vol.117, pp.945–71.

Baylis, G.C., Driver, J., Baylis, L., and Rafal, R.D. (1994) 'Reading of letters and words in a patient with Balint's syndrome', *Neuropsychological*, vol.32, pp.1273–86.

Bisiach, E. and Luzzatti, C. (1978) 'Unilateral neglect of representational space', *Cortex*, vol.14, pp.129–33.

Broadbent, D.E. (1952) 'Listening to one of two synchronous messages', *Journal of Experimental Psychology*, vol.44, pp.51–5.

Broadbent, D.E. (1954) 'The role of auditory localization in attention and memory span', *Journal of Experimental Psychology*, vol.47, pp.191–6.

Broadbent, D.E. and Broadbent, M.H.P. (1987) 'From detection to identification: response to multiple targets in rapid serial visual presentation', *Perception and Psychophysics*, vol.42, pp.105–13.

Broadbent, D.E. and Gathercole, S.E. (1990) 'The processing of non-target words: semantic or not?' *Quarterly Journal of Experimental Psychology*, vol.42A, pp.3–37.

Buchner, A., Irmen, L., and Erdfelder, E. (1996) 'On the irrelevance of semantic information for the "Irrelevant Speech" effect', *Quarterly Journal of Experimental Psychology*, vol.49A, pp.765–79.

Caligiuri, M.P., Hellige, J.B., Cherry, B.J., Kwok, W., Lulow, L.L., and Lohr, J.B. (2005) 'Lateralized cognitive dysfunction and psychotic symptoms in schizophrenia', *Schizophrenia Research*, vol.80, pp.151–61.

Cheesman, J. and Merikle, P.M. (1984) 'Priming with and without awareness', *Perception and Psychophysics*, vol.36, pp.387–95.

Coltheart, M. (1980) 'Iconic memory and visible persistence', *Perception and Psychophysics*, vol.27, pp.183–228.

Corteen, R.S. and Wood, B. (1972) 'Autonomous responses to shock-associated words in an unattended channel', *Journal of Experimental Psychology*, vol.94, pp.308–313.

De Pascalis, V. (2007) 'Phase-ordered gamma oscillations and the modulation of hypnotic experience', in Jamieson, G. (ed.) *Hypnosis and Conscious States: The Cognitive-Neuroscience Perspective*, Oxford, Oxford University Press.

Deutsch, J.A. and Deutsch, D. (1963) 'Attention: some theoretical considerations', *Psychological Review*, vol.70, pp.80–90.

Driver, J. (1996) 'Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading', *Nature*, vol.381, pp.66–8.

Driver, J. and Halligan, P.W. (1991) 'Can visual neglect operate in object centred co-ordinates? An affirmative case study', *Cognitive Neuropsychology*, vol.8, pp.475–96.

Dudley, R., O'Brien, J., Barnett, N., McGuckin, L., and Britton, P. (2002) 'Distinguishing depression from dementia in later life: a pilot study employing the emotional Stroop task', *International Journal of Geriatric Psychiatry*, vol.17, pp.48–53.

Duncan, J. and Humphreys, G.W. (1989) 'Visual search and visual similarity', *Psychological Review*, vol.96, pp.433–58.

Evett, L.J. and Humphreys, G.W. (1981) 'The use of abstract graphemic information in lexical access', *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, vol.33A, pp.325–50.

Gellatly, A., Cole, G., and Blurton, A. (1999) 'Do equiluminant object onsets capture visual attention?' *Journal of Experimental Psychology: Human Perception and Performance*, vol.25, pp.1609–24.

Giesbrecht, B. and Di Lollo, V. (1998) 'Beyond the attentional blink: visual masking by object substitution', *Journal of Experimental Psychology: Human Perception and Performance*, vol.24, pp.1454–66.

Gronau, N., Cohen, A., and Ben-Shakhar, G. (2003) 'Dissociations of personally significant and task-relevant distractors inside and outside the focus of attention: a combined behavioral and psychophysiological study', *Journal of Experimental Psychology: General*, vol.132, pp.512–29.

Gronau, N., Cohen, A., and Ben-Shakhar, G. (2009) 'Distractor interference in focused attention tasks is not mediated by attention capture', *Quarterly Journal of Experimental Psychology*, vol.62, pp.1685–95.

Haig, A.R., Gordon, E., De Pascalis, V., Meares, R.A., Bahramali, H., and Harris, A. (2000) 'Gamma activity in schizophrenia: evidence of impaired network binding?' *Clinical Neurophysiology*, vol.111, pp.1461–8.

Humphreys, G.W. (2001) 'A multi-stage account of binding in vision: neuropsychological evidence', *Visual Cognition*, vol.8, pp.381–410.

Humphreys, G.W., Riddoch, M.J., Forti, S., and Ackroyd, K. (2004) 'Action influences spatial perception: neuropsychological evidence', *Visual Cognition*, vol.11, pp.401–27.

Humphreys, G.W., Wulff, M., Yoon, E.Y., and Riddoch, M.J. (2010) 'Neuropsychological evidence for visual- and motor-based affordance: effects of reference frame and object–hand congruence', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.36, pp.659–70.

Jones, D.M. (1999) 'The cognitive psychology of auditory distraction', *British Journal of Psychology*, vol.90, pp.167–87.

Jones, D.M. and Macken, W.J. (1995) 'Phonological similarity in the irrelevant speech effect: within- or between-stream similarity?' *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.21, pp.103–15.

Jones, D.M., Miles, C., and Page, J. (1990) 'Disruption of reading by irrelevant speech: effects of attention, arousal or memory?' *Applied Cognitive Psychology*, vol.4, pp.89–108.

Jones, D.M., Saint-Aubin, J., and Tremblay, S. (1999) 'Modulation of the irrelevant sound effect by organizational factors: further evidence from streaming by location', *Quarterly Journal of Psychology*, vol.52A, pp.545–54.

Kolers, P.A. (1968) 'Some psychological aspects of pattern recognition', in Kolers, P.A. and Eden, M. (eds.) *Recognizing Patterns*, Cambridge, MA, MIT Press.

Kosslyn, S.M., Thompson, W.L., Costantini-Ferrando, M.F., Alpert, N.M., and Spiegel, D. (2000) 'Hypnotic visual illusion alters color processing in the brain', *American Journal of Psychiatry*, vol.157, pp.1279–84.

Kouider, S. and Dehaene, S. (2007) 'Levels of processing during non-conscious perception: a critical review of visual masking', *Philosophical Transactions of the Royal Society, B*, vol.362, pp.857–75.

Lamme, V.A.F. (2003) 'Why visual attention and awareness are different', *Trends in Cognitive Sciences*, vol.7, pp.12–18.

Merikle, P.M. and Joordens, S. (1997) 'Parallels between perception without attention and perception without awareness', *Consciousness and Cognition*, vol.6, pp.219–36.

Naish, P.L.N. (1990) 'Simulating directionality in airborne auditory warnings and messages', in Life, M.A., Narborough-Hall, C.S., and Hamilton, W.I. (eds.) *Simulation and the User Interface*, London and New York, Taylor and Francis.

Naish, P.L.N. (2007) 'Time distortion, and the nature of hypnosis and consciousness', in Jamieson, G. (ed.) *Hypnosis and Conscious States: The Cognitive Neuroscience Perspective*, Oxford, Oxford University Press.

Naish, P.L.N. (2010) 'Hypnosis and hemispheric asymmetry', *Consciousness and Cognition*, vol.19, pp.230–4.

Norman, D.A. (1968) 'Towards a theory of memory and attention', *Psychological Review*, vol.75, pp.522–36.

Pecher, D., Zeelenberg, R., and Raaijmakers, G.W. (2002) 'Associative priming in a masked perceptual identification task: evidence for automatic processes', *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, vol.55A, pp.1157–73.

Raymond, J.E., Shapiro, K.L., and Arnell, K.A. (1992) 'Temporary suppression of visual processing in an RSVP task: an attentional blink?' *Journal of Experimental Psychology: Human Perception and Performance*, vol.18, pp.849–60.

Riddoch, M.J., Humphreys, G.W., Edwards, S., Baker, T., and Willson, K. (2003) 'Actions glue objects but associations glue words: neuropsychological evidence for multiple object selection', *Nature Neuroscience*, vol.6, pp.82–9.

Shaffer, W.O. and LaBerge, D. (1979) 'Automatic semantic processing of unattended words', *Journal of Verbal Learning and Verbal Behaviour*, vol.18, pp.413–26.

Shapiro, K.L., Caldwell, J., and Sorensen, R.E. (1997) 'Personal names and the attentional blink: a visual "cocktail party" effect', *Journal of Experimental Psychology: Human Perception and Performance*, vol.23, pp.504–514.

Spence, C. (2002) 'Multisensory integration, attention and perception', in Roberts, D. (ed.) *Signals and Perception: The Fundamentals of Human Sensation*, Basingstoke, Hampshire/New York, Palgrave Macmillan.

Sperling, G. (1960) 'The information available in brief visual presentations', *Psychological Monographs*, vol.74, pp. 1–29.

Stroop, J.R. (1935) 'Studies of interference in serial verbal reactions', *Journal of Experimental Psychology*, vol.18, pp.643–62.

Szechtman, H., Woody, E., Bowers, K.S., and Nahmias, C. (1998) 'Where the imaginal appears real: a positron emission tomography study of auditory hallucinations', *Proceedings of the National Academy of Sciences*, vol.95, pp.1956–60.

Treisman, A. (1960) 'Contextual cues in selective listening', *Quarterly Journal of Experimental Psychology*, vol.12, pp.242–8.

Treisman, A. (1998) 'Feature binding, attention and object perception', *Philosophical Transactions of the Royal Society of London, B*, vol.353, pp.1295–306.

Treisman, A. and Gelade, G. (1980) 'A feature-integration theory of attention', *Cognitive Psychology*, vol.12, pp.97–136.

Turvey, M.T. (1973) 'On peripheral and central processes in vision: inferences from an information-processing analysis of masking with patterned stimuli', *Psychological Review*, vol.80, pp.1–52.

Vogel, E.K., Luck, S.J., and Shapiro, K.L. (1998) 'Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink', *Journal of Experimental Psychology: Human Perception and Performance*, vol.24, pp.1656–74.

Woldorff, M.G., Gallen, C.C., Hampson, S.A., Hillyard, S.A., Pantev, C., Sobel, D., and Bloom, F.E. (1993) 'Modulation of early sensory processing in human auditory cortex during auditory selective attention', *Proceedings of the National Academy of Sciences of the USA*, vol.90, pp.8722–6.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

CHAPTER 3

PERCEPTION

Graham Pike, Graham Edgar, and Helen Edgar

1 INTRODUCTION

If you have ever searched frantically for an object that turns out to have been right in front of you all along, then this chapter may make you feel better. For, as you will see, perception of even the simplest object is actually a very complex affair. So, next time you turn the house upside down looking for your keys and then find them in the first place you looked, remember that your brain is using extremely sophisticated processes, many of which are beyond even the most advanced computer programs available today (not that computer programs ever lose their keys!).

The sophistication of the cognitive processes that allow us to perceive visually is perhaps, if perversely, revealed best through the errors that our perceptual system can make. Figure 3.1 contains three very simple visual illusions that illustrate this. Image (a) is the Müller–Lyer illusion, in which the vertical line on the left is perceived as being longer even though both lines are of an identical length. Image (b) is a Necker cube, in which it is possible to perceive the cube in either of two perspectives (although you can never see both at the same time so please do not strain your eyes trying). Image (c) is Kanizsa's (1976) illusory square, in which a square is perceived even though the image does not contain a square but only four three-quarter-complete circles.

If the cognitive processes involved in perception were simple, then it would be hard to see how the effects in Figure 3.1 could occur. After all, they are all based on very straightforward geometric shapes that should be easy to perceive accurately. Activity 3.1 demonstrates that there must be more sophisticated processes that have been developed to perceive the complex visual environment, which get confused or tricked by elements of these images. In fact the three effects above are likely to be caused because our visual system has evolved to perceive solid, three-dimensional (3D) objects and attempts to interpret the two-dimensional (2D) shapes as resulting from 3D scenes.

Perceptual errors arising from localized damage to the brain also demonstrate the complexities involved in visual perception. Some of the problems faced by people suffering from specific neuropsychological conditions include: being able to recognize objects but not faces (prosopagnosia); being able to perceive individual parts of the environment but not to integrate these parts into a whole; believing that one's family has been replaced by robots/aliens or impostors of the same appearance (Capgras syndrome); and only being able to perceive one side of an object (visual, or sensory, neglect – see Chapter 2, Section 5.1).

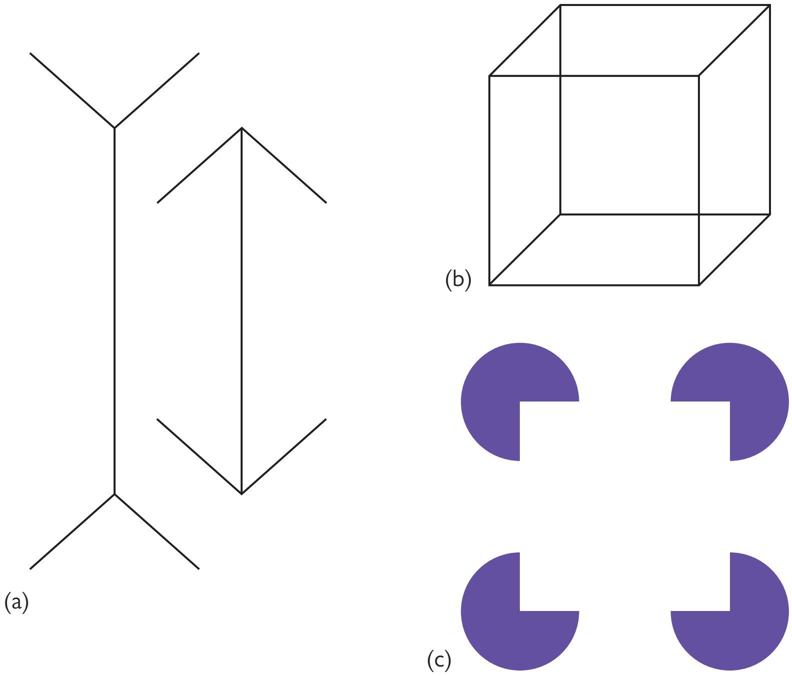


FIGURE 3.1 Three visual phenomena: (a) Müller–Lyer illusion; (b) Necker cube; (c) Kanizsa's illusory square.

✱ ACTIVITY 3.1

Look at each of the three visual illusions in Figure 3.1 and try to work out why it occurs. If you can't think of an answer, it may help to look at Figure 3.2.

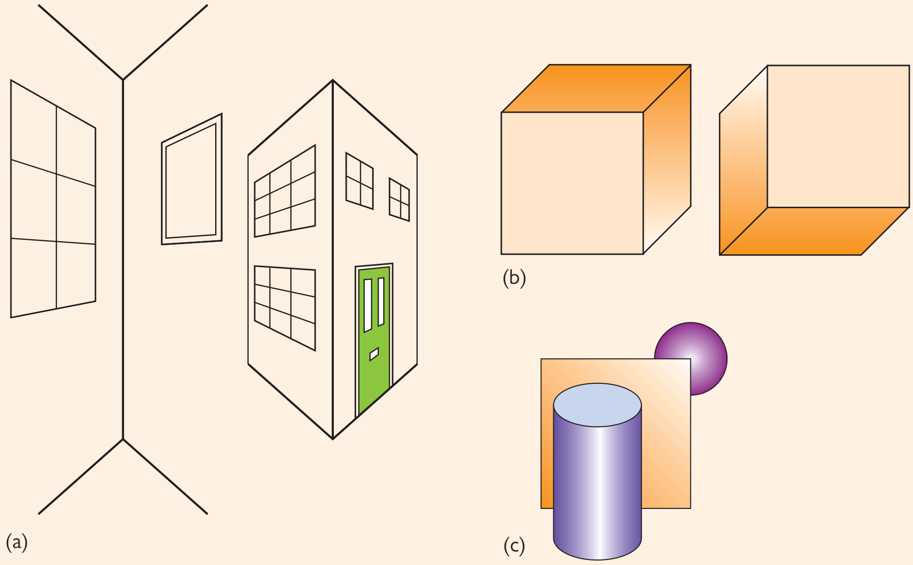


FIGURE 3.2 Some clues as to why the illusions in Figure 3.1 may occur.

COMMENT

One explanation for the Müller–Lyer illusion is that the arrowheads provide clues as to the distance of the upright line. For example, the inward-pointing arrowheads suggest that the vertical line might be the far corner of a room, whilst the outward-pointing arrowheads suggest that the vertical line could be the near corner of a building. We therefore see the first vertical line as longer because we assume it is further away from us than the second vertical line, though it makes the same size image on the retina.

The Necker cube can be seen in two different ways, as there are no clues as to which is the nearest face. Most cube-like objects that we encounter are solid and contain cues from lighting and texture about which is the nearest face. As the Necker cube does not contain these cues, we are unable to say for certain which face is closest.

Kanizsa's illusory square occurs due to a phenomenon known as perceptual completion. When we see an object partly hidden behind (occluded by) another object, we represent it to ourselves as a whole object rather than as missing its hidden parts. In the same way, we assume that four black circles are being occluded by a white square.

1.1 Perceiving and sensing

The term perception has different meanings, although a common element in most meanings is that perception involves the analysis of sensory information. When cognitive psychologists talk about perception, they are usually referring to the basic cognitive processes that analyse information from the senses. Throughout this chapter we shall be examining research and theories that have attempted to reveal and describe the cognitive processes responsible for analysing sensory information and providing a basic description of our environment; basically, how we make sense of our senses!

There has been considerable debate about the role played by sensory information in our perception of the world, with some philosophers rejecting the idea that it plays any part at all in the perception of objects. Atherton (2002) suggested that this may be because the notion of a sensation is rather problematic: 'Sensations seem to be annoying, extra little entities . . . that somehow intervene between the round dish and our perception of it as round' (Atherton, 2002, p.4). We will not delve into this philosophical debate here, other than to note the distinction between sensation and perception. Throughout this chapter we will use the term 'sensation' to refer to the ability of our sense organs to detect various forms of energy (such as light or sound waves). However, to sense information does not entail making sense of it. There is a key difference between being able to detect the presence of a certain type of energy and being able to make use of that energy to provide information as to the nature of the environment surrounding us. Thus we use the term 'sensation' to refer to that initial detection and the term 'perception' to refer to the process of constructing a description of the surrounding world. For example, there is a difference between the cells in a person's eye reacting to light (sensation) and that person knowing that their course tutor is offering them a cup of tea (perception).

You may have noticed that we have begun to focus on visual perception rather than any of the other senses. Although the other senses, particularly hearing and touch, are undoubtedly important, there has been far more research on vision than on the other modalities. This is because when we interact with the world we rely more on vision than on our other senses. As a result, far more of the primate brain is engaged in processing visual information than in processing information from any of the other senses. We use vision in both quite basic ways, such as avoiding objects, and in more advanced ways, such as in reading or recognizing faces and objects. So, although the previous chapter examined auditory perception and Chapter 4 will explore haptic perception (touch) as

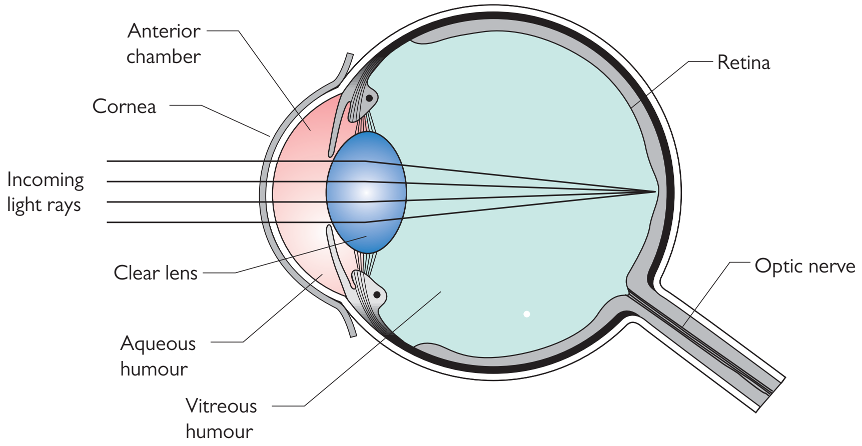


FIGURE 3.3 The human eye.

well as visual perception, we will devote the present chapter to examining research into, and theories of, visual perception.

1.2 The eye

The logical place to start any consideration of visual perception is with the eye. A cross-section of the human eye is presented in Figure 3.3. Incoming light passes through the cornea into a small compartment called the anterior chamber (filled with fluid termed aqueous humour) and then through the lens into the major chamber of the eye that is filled with a viscous jelly called vitreous humour. The light is focused by the lens/cornea combination onto the retina on the back surface of the eye. It is the receptor cells in the retina that 'sense' the light.

The retina consists of two broad classes of receptor cell, rods and cones (so called for their shapes). Both rods and cones are sensitive to light, although the rods respond better than the cones at low light levels and are therefore the cells responsible for maintaining some vision in poor light. The cones are responsible for our ability to detect fine detail and different colours and are the basis of our vision at higher (daylight) light levels. Many animals, such as dogs and cats, have a higher ratio of rods to cones than humans do. This allows them to see better in poor light, but means that they are not so good at seeing either colour or fine detail.

One area of the retina that is of particular interest is the central portion known as the macula lutea (it is yellow in colour and 'lutea' derives from a Latin word that means yellow), which contains almost all of the cones within the human retina. Within the macula, there is a small indentation called the fovea. The fovea is the area of the retina that contains the highest density of cones and is responsible for the perception of fine detail.

✱ ACTIVITY 3.2

Place your thumbs together and hold them out at arm's length from your eyes. Now slowly move your left thumb to the left whilst keeping your eyes focused on your right thumb. You will find that after you have moved your left thumb more than about two thumb widths away from where your eyes are focused, it appears to go out of focus. This is because the light being reflected into your eyes from the left thumb is no longer striking the fovea, meaning that you cannot perceive it in fine detail.

1.3 Approaches to perception

Psychologists have taken many different approaches to studying perception. One important distinction between approaches is whether the 'goal' of perception is assumed to be action or recognition. It is possible to conceive of recognition and action as being stages in the same perceptual process, so that action would only happen once recognition had taken place. However, our reaction to objects in the environment sometimes has to be very quick indeed, so that first having to work out what an object may be would be inconvenient to say the least. For example, if I see a moving object on a trajectory that means it will hit me in the head, then the most important thing is to move my head out of the way. Working out that the object is the crystal tumbler containing vodka and tonic that was only moments ago in the hand of my somewhat angry-looking partner is, for the moment at least, of secondary importance. I need to act to get out of the way of the object regardless of what the object actually is or who threw it.

As we shall see, there is evidence that perception for action and perception for recognition are quite different processes that may involve different neural mechanisms (Milner and Goodale, 1998). But, although it is important to make the distinction between perception for action and perception for recognition, we should not see them as being entirely independent. Sometimes the object that is about to hit your head could be the football that David Beckham has just crossed from the wing, requiring a very different response from that to the crystal tumbler.

Another way of differentiating approaches to perception is to consider the 'flow of information' through the perceptual system. To see what we mean by this phrase, try Activity 3.3.

So, in the case of touch, perception of the environment can involve information 'flowing' through the relevant perceptual system in two directions. But what about vision? If we were to remove the blindfold from our student in Activity 3.3, they would instantly be able to tell what the unknown object was or to spot the book in front of them. Does this mean that there is not a similar flow of information when the sense being used is vision?

To answer this question, let's try to formulate the stages involved in the student perceiving that there is a book in front of them. One approach might be:

- Light reflected from the book strikes the retina and is analysed by the brain.

✱ ACTIVITY 3.3

Consider these two scenarios:

1 A blindfolded student trying to work out what the unknown object they have been handed might be.

2 A blindfolded student searching for their textbook.

Imagine you are the blindfolded student. What strategies do you think you might employ to complete the above two tasks successfully? Can you identify any key differences in these strategies?

COMMENT

A common strategy to employ for the first scenario is to try to build up a 'picture' of the object by gradually feeling it. A common strategy to employ for the second scenario is to hold in your mind the likely shape and texture of the book and to search the environment for an object that shares these characteristics. The key difference between these scenarios is the direction in which information about the object is 'flowing', demonstrated by how the student's existing knowledge of what objects look like is being utilized. In the first scenario, information is flowing 'upward', starting with an analysis of the information derived from the senses (in this case via touch). In the second scenario, information is flowing 'downward', starting with the knowledge of what books tend to feel like.

- This analysis reveals four sudden changes in brightness (caused by the edges of the book against whatever is behind it).
- Two of these are vertical edges and two are horizontal edges (the left/right and top/bottom of the book).
- Each straight edge is joined (by a right angle at each end) to two others (to form the outline of the book).
- Within these edges is an area of gradually changing brightness containing many small, much darker areas (the white pages with a growing shadow toward the spine and the much darker words).
- A comparison of this image with representations of objects seen previously suggests that the object is an open book.

This approach starts with the image formed on the retina by the light entering the eye and proceeds by analysing this pattern to gradually build up a representation of the object in view, so we refer to it as involving **bottom-up processing**. This means that the flow of information through the perceptual system starts from the bottom – the sensory receptors – and works upward until an internal representation of the object is formed.

There is, however, another way of recognizing the book. It is very likely that the student has seen many books in the past and has a fair idea of what a book should look like. This existing knowledge regarding book appearance could come in very useful in finding the textbook. Instead of building up a picture of the environment by analysing sensory information alone, it could be that the student uses existing knowledge of what books look like to find this particular book. For example, they might progress like this:

- I know that books are rectangular in shape and have light pages with dark words.
- I can see something in front of me that matches this description, so it must be a book.

The flow of information in this latter example has been reversed. The student started with existing knowledge regarding the environment and used this to guide their processing of sensory information. Thus the flow of information progressed from the top down, as it started with existing knowledge stored in the brain, and we refer to it as involving **top-down processing**.

So both haptic and visual perceptual processes may operate both by building up a picture of the environment from sensory information and by using existing knowledge to make sense of new information. In other words, the flow of information through the perceptual system can be either bottom-up or top-down. These concepts will be explored throughout this chapter and we shall examine theories that concentrate on one or the other of these processes, and also look at how they might interact.

SUMMARY OF SECTION 1

- Even the perception of simple images involves sophisticated cognitive processing, as demonstrated by visual illusions and neuropsychological disorders.
- We use the term sensation to refer to the detection of a particular form of energy by one of the senses and the term perception to refer to the process of making sense of the information sent by the senses.
- In the human eye the lens and cornea focus light onto the retina, which contains receptor cells that are sensitive to light.
- Perception can have different goals. The most common goals are perception for action and perception for recognition.
- The bottom-up approach to perception sees sensory information as the starting point, with perception occurring through the analysis of this information to generate an internal description of the environment.
- The top-down approach to perception involves making greater use of prior knowledge, with this guiding the perceptual process.

2 THE GESTALT APPROACH TO PERCEPTION

As with Chapter 2, we are going to examine the various approaches that have been taken to studying visual perception in a more or less historical order. One of the principal approaches to perception in the first half of the twentieth century was that of the **Gestalt movement**, which was guided by the premise 'the whole is greater than the sum of its parts'. In perceptual terms, this meant that an image tended to be perceived according to the organization of the elements within it, rather than according to the nature of the individual elements themselves.

It is easy to see **perceptual organization** at work as it tends to be a very powerful phenomenon. In fact, it appears as if both visual and auditory stimuli can be grouped according to similar organizing principles (Aksentijevic *et al.*, 2001).

Have a look at Activity 3.4. Most people looking at these images describe a circle and two crossing lines. But the image on the left is not a circle as it contains a gap at the top. This is the Gestalt perceptual organizational phenomenon of closure at work, in which a 'closed' figure tends to be perceived rather than an 'open' one. Likewise, the image on the right is not necessarily showing crossing lines, as it could be two pen-tips touching (in the middle of the image). The reason you see a cross is due to what the Gestalt researchers called good continuation, by which we tend to interpret (or organize) images to produce smooth continuities rather than abrupt changes.

Try Activity 3.5. You will probably have seen these images as two squares, due to the law of closure. However, you will also probably have seen the square on the left as consisting of columns of dots and the one on the right as consisting of rows of dots. If so, this was an example of the organizational law of proximity, because in the left-hand image the horizontal spacing between the dots is greater than the vertical, and vice versa for the image on the right. Thus, the proximity of the individual elements is being used to group them into columns in the left-hand square and rows in the right-hand one.

As well as again seeing a square due to the law of closure. Now try Activity 3.6. Again, you probably saw this as a square, due to the law of closure, but perhaps also saw the square as consisting of columns of circles. If so, this was an example of the organizational law of similarity (in this case the similarity of colour). However, the spacing of the circles is such that the law of proximity encourages you to see rows

✱ ACTIVITY 3.4

Look at Figure 3.4 and describe your first impression of what you see.

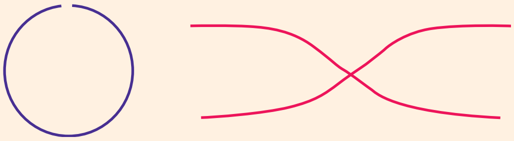


FIGURE 3.4 Two examples of perceptual organization.

✱ ACTIVITY 3.5

As before, look at Figure 3.5 and describe your first impressions.

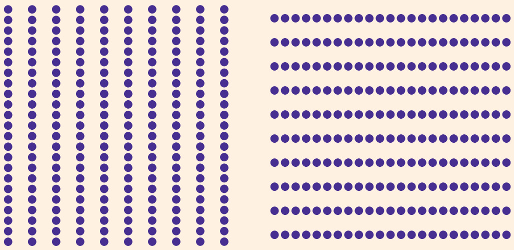


FIGURE 3.5 The organizational law of proximity.

✱ ACTIVITY 3.6

Now describe what you see in Figure 3.6.

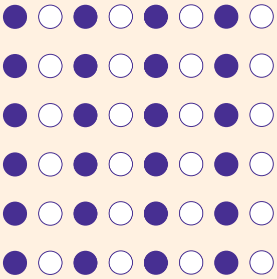


FIGURE 3.6 The organizational laws of similarity and proximity.

not columns. For many people the law of similarity takes precedence and they see columns, while others may tend to see rows. Most people can readily switch between one organization (or interpretation) and the other because each conforms with a particular Gestalt law.

The Gestalt researchers (including Koffka, 1935, Kohler, 1947, and Werthiemer, 1923) formulated other organizational laws, but most were deemed to be manifestations of the **Law of Pragnanz**, described by Koffka as: 'Of several geometrically possible organizations that one will actually occur which possesses the best, simplest and most stable shape' (Koffka, 1935, p.138).

So, you can see that a number of organizational laws can be used in order to work out which individual components of an image should be grouped together. Now look around the room in which you are sitting. How many squares composed of dots can you see? How many nearly complete circles and crossing lines are there? Your immediate response was probably to say 'none' or 'only those in this book'. However, if you look carefully you will see that the stimuli used in the Gestalt demonstrations do have counterparts in the real world. For example, when I look out of my window I see a football that is partly hidden by a post, which provides an example of closure, as I perceive a complete sphere rather than an incomplete circle. The figures that you have seen in this section can therefore be seen as simplified 2D versions of real-world objects and scenes. Because they are simplified, some information that would be present in real-world scenes is discarded. This lack of realism is a disadvantage. On the other hand, however, it is possible to control and manipulate features of these figures, such as the proximity or similarity of elements, to see how they may contribute to perception.

As we shall see in the next section, there is considerable tension in the field of visual perception regarding the usefulness of simplified stimuli such as those used by the Gestaltists. Some approaches are based on laboratory experimentation in which simplified scenes or objects are shown to participants, whilst proponents of other approaches claim that perception can only be studied in the real world, by examining how people perceive solid, 3D objects that are part of a complex 3D environment.

SUMMARY OF SECTION 2

- The Gestalt approach to perception involved studying the principles by which individual elements tend to be organized together.
- Organizing principles include closure, good continuation, proximity, and similarity.
- The stimuli used by Gestalt researchers tended to be quite simple, two-dimensional geometric patterns.

3 GIBSON'S THEORY OF PERCEPTION

In Section 1.3 we stated that one way of classifying different approaches to perception was according to whether they were primarily bottom-up or top-down. If visual perception is based primarily around bottom-up processing, we must be capable of taking the information from the light waves that reach our eyes and refining it into a description of the visual environment. Bottom-up perception requires that the light arriving at the retina is rich in information about the environment. One bottom-up approach to perception, that of J.J. Gibson (1950, 1966), is based on the premise that the information available from the visual environment is so rich that no cognitive processing is required at all. As Gibson himself said:

> *When the senses are considered as a perceptual system, all theories of perception become at one stroke unnecessary. It is no longer a question of how the mind operates on the deliverances of sense, or how past experience can organize the data, or even how the brain can process the inputs of the nerves, but simply how information is picked up.*
>
> *(Gibson, 1966, p.319)*

If you are thinking to yourself, 'What does picked up mean?' or 'How is this information *picked up*?', you are expressing a criticism that is often levelled at Gibson's theory (e.g. Marr, 1982). The Gibsonian approach concentrates on the information present in the visual environment rather than on how it may be analysed. There is a strong link between perception and action in Gibson's theory, and action rather than the formation of an internal description of the environment can be seen as the 'end point' of perception.

Gibson conceptualized the link between perception and action by suggesting that perception is **direct**, in that the information present in light is sufficient to allow a person to move through and interact with the environment. One implication of this is that, whereas perception of a real environment is direct, perception of a 2D image in a laboratory experiment (or any 2D image come to that) would be *indirect*. When confronted with an image, our direct perception is that it is an image; that it is two-dimensional and printed on paper, for example. Our perception of whatever is being depicted by the image is only indirect. For this reason, Gibson thought that perception could never be fully explored using laboratory experiments.

When you look at Figure 3.7, what do you see? Your first reaction is probably to say 'a pipe'. But if what you are seeing is a pipe, then why can't you pick it up and smoke it? As Magritte informs us, what you are seeing is not a pipe but a picture of a pipe. Like Gibson, Magritte is drawing a distinction between direct perception (paint on canvas) and indirect perception (that the painting depicts a pipe).

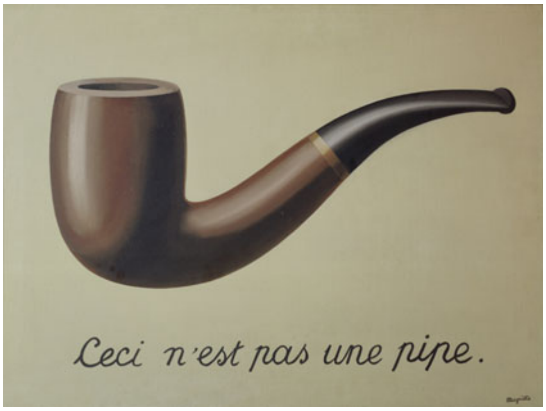

Ceci n'est pas une pipe.
Magritte

FIGURE 3.7 *Ceci n'est pas une pipe*, 1928, by René Magritte.

3.1 An ecological approach

At the heart of Gibson's approach to perception is the idea that the world around us structures the light that reaches the retina. Gibson believed perception should be studied by determining how the real environment structures the light that reaches our retina. From the importance placed on the 'real world' it is clear why Gibson called his theory of perception ecological optics. Gibson referred to theories that were based on experiments employing artificial, isolated, flat (or plane) shapes as 'air' theories, whilst he referred to his own as a 'ground' theory, as it emphasized the role played by the real, textured surface of the ground in providing information about distance. As Gibson stated: 'A surface is substantial; a plane is not. A surface is textured; a plane is not. A surface is never perfectly transparent; a plane is. A surface can be seen; a plane can only be visualized' (Gibson, 1979, p.35).

The impetus for Gibson's theory came from his work training pilots to land and take off during the Second World War. When approaching a runway, it is very important that a pilot is able to judge accurately the distance between the plane and the ground. The perceptual skill involved in this judgement is that of 'depth perception', this being the ability to judge how far you are from an object or surface. However, Gibson found that tests based on pictorial stimuli did not distinguish good from bad pilots and that training with pictorial stimuli had little impact on actual landing performance (Gibson, 1947). Extrapolating from this problem, Gibson suggested that psychological experimentation based on the use of pictorial stimuli is not an apt method for studying perception.

His point was that the experience of perception in the real world is very different from the experience of looking at 2D experimental stimuli in a laboratory. In the real world, objects are not set against a blank background but against the ground, which consists of a very large number of surfaces that vary in their distance from and orientation to the observer. In their turn, these surfaces are not perfectly smooth planes but consist of smaller elements, such as sand, earth, and stone, which give them a textured appearance. In addition, the objects themselves will consist of real surfaces that also contain texture. To explain perception, we need to be able to explain how these surfaces and textures provide information about the world around us.

3.2 The optic array and invariant information

The structure that is imposed on light reflected by the textured surfaces in the world around us is what Gibson termed the ambient optic array. The basic structure of the optic array is that the light reflected from surfaces in the environment converges at the point in space occupied by the observer (see Figure 3.8). As you can see from Figure 3.9, as you stand up, the position of your head with respect to the environment is altered and the optic array changes accordingly.

You can see from Figures 3.8 and 3.9 that the primary structure of the optic array is a series of angles that are formed by light reflecting into the eyes from the surfaces within the environment. For example, an angle may be formed between the light that is reflected from the near edge of a table and that from the far edge.

In addition to the primary structure of the optic array, Gibson maintained that there were additional, higher-order features that could provide unambiguous information as to the nature of the environment. He referred to these higher-order features as invariants and believed that an observer could perceive the surrounding world by actively sampling the optic array in order to detect invariant information.

One of the most commonly cited forms of invariant information was explored by Sedgwick (1973). Sedgwick demonstrated the 'horizon ratio relation', which specifies that the ratio of how much of an object is above the horizon to how much is below it remains constant (or invariant) as the object travels either toward or away from you (see Figure 3.10). This form of invariant information allows you to judge the relative heights of different objects regardless of how far away they are. The proportion of the object that is

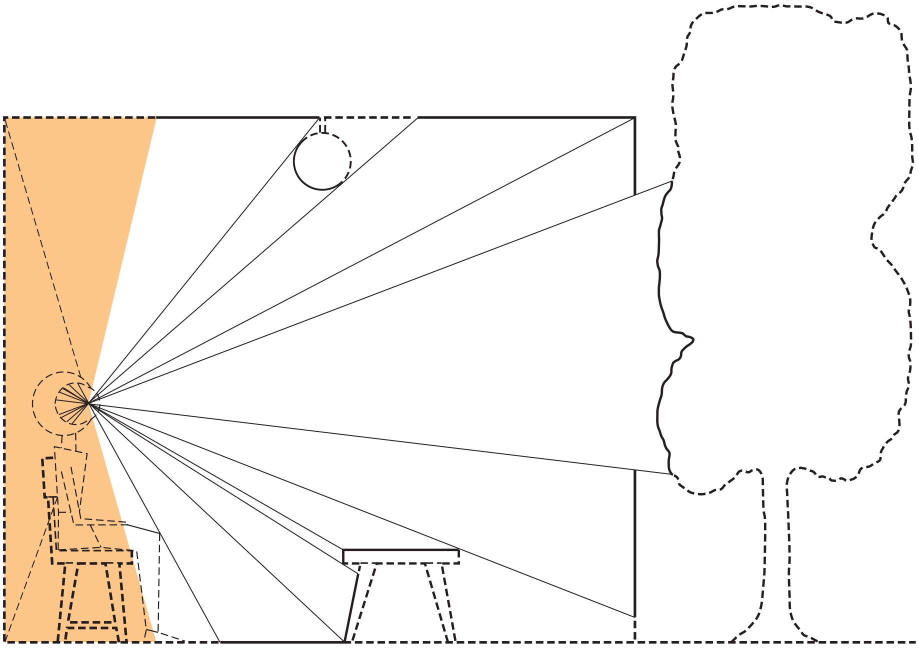


FIGURE 3.8 The ambient optic array. *Source: Gibson, 1979, Figure 5.3*

'above' the horizon increases with the overall height of the object (see Figure 3.11).

One of the most important forms of invariant information in Gibson's theory is **texture gradient**, although he also discusses gradients of colour, intensity, and disparity. There are three main forms of texture gradient relating to the density, perspective, and compression of texture elements. The exact nature of a texture element will change from surface to surface (see Figure 3.12); in a carpet the elements are caused by the

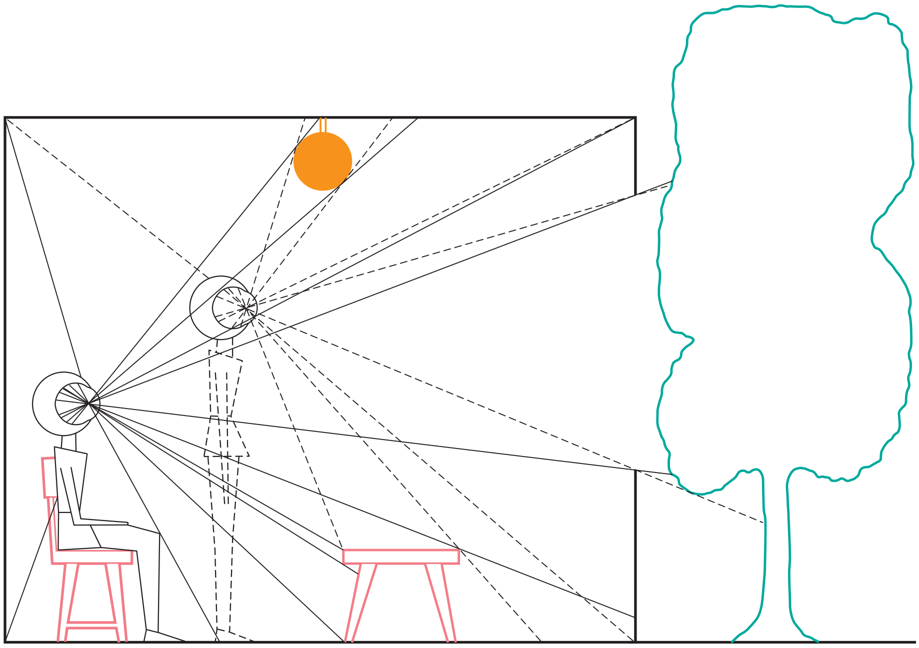


FIGURE 3.9 Change in the optic array caused by movement of the observer. *Source: Gibson, 1979, Figure 5.4*

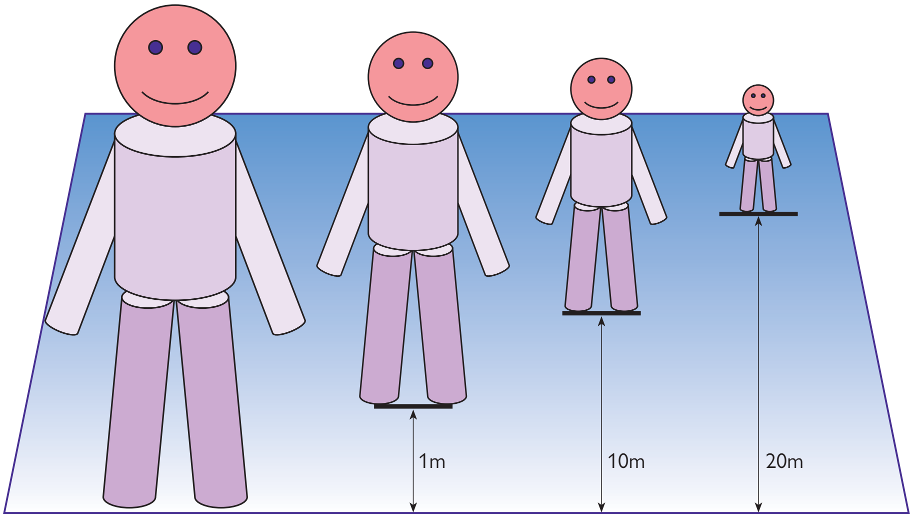


FIGURE 3.10 The horizon ratio relation: same height objects at different distances.

individual twists of material, while on a road they are caused by the small stones that make up the surface. In making use of texture gradients, we assume that the texture of the surface is uniform; for example, that the road surface consists of stones of similar size throughout its length. Therefore, any change in the apparent nature of the texture provides us with information regarding the distance, orientation, and curvature of the surface.

Using texture gradients as a guide, we can tell if a surface is receding because the density of texture elements (number of elements per square metre) will increase with distance. For example, the surface in Figure 3.13(a) appears to recede as the density of texture elements (the individual squares) increases toward the top of the image.

In a similar fashion, the perspective gradient (the width of individual elements) and the compression gradient (the height of individual elements) can reveal the shape and orientation of a surface. As you can see from Figure 3.13(b), we do not see this surface as flat because the width and height of the individual texture

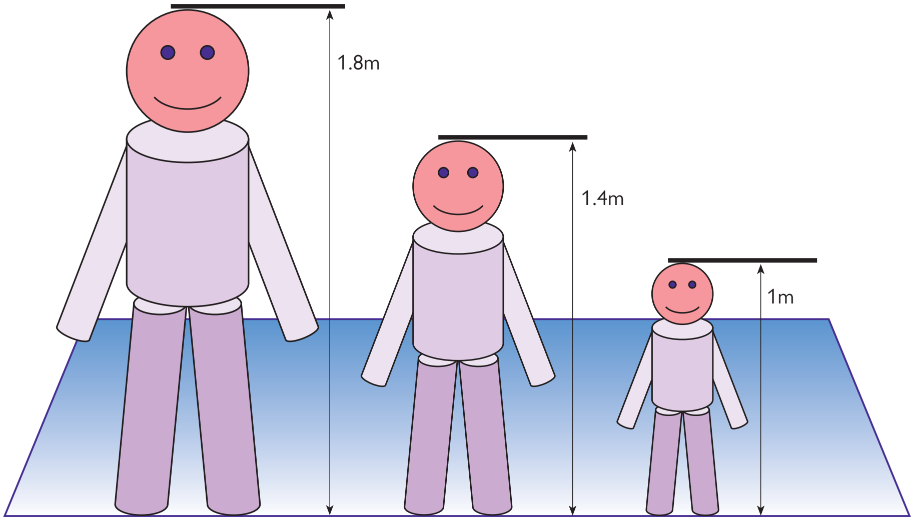


FIGURE 3.11 The horizon ratio relation: different height objects at same distance.

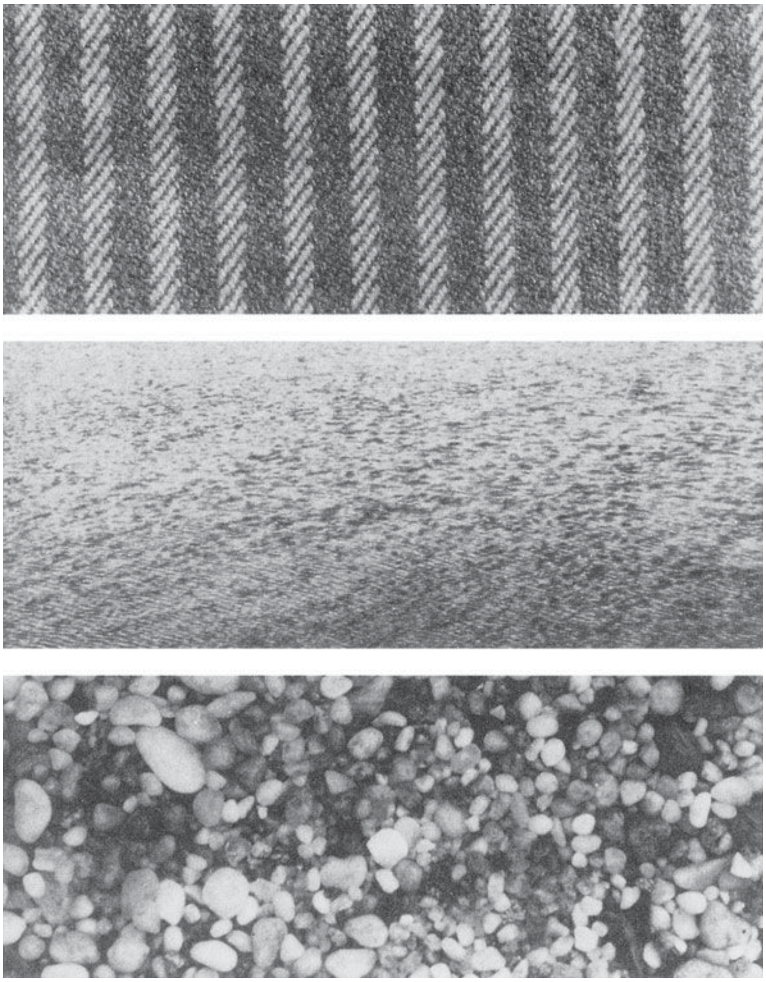
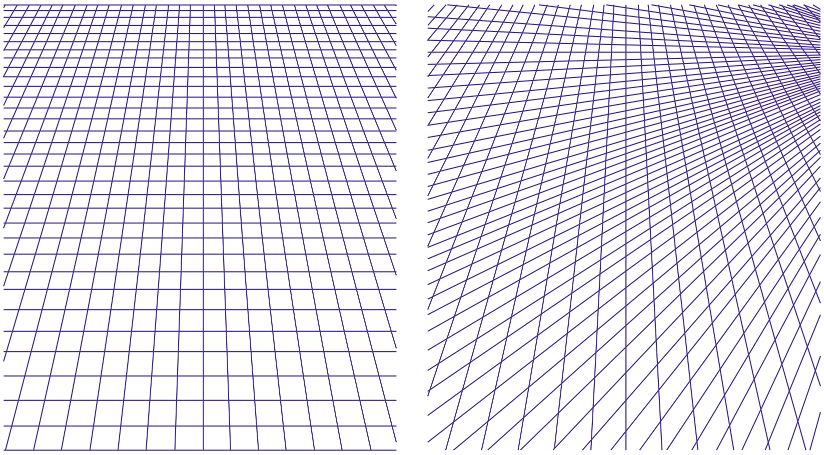


FIGURE 3.12 Examples of texture elements. *Source: Gibson, 1979, Figure 2.1*



(a) (b)

FIGURE 3.13 (a) How texture gradient can reveal that a surface is receding; (b) How perspective and compression gradients reveal the shape and orientation of a surface.

elements changes, making the surface appear to be slanting and curved.

Without texture, considerable ambiguity about shape and orientation can be introduced into the stimulus and this poses a problem for experiments that make use of planar geometric shapes (as you saw with the Necker cube in Activity 3.1). So, texture gradient is a powerful source of invariant information provided by the structure of light within the optic array. It furnishes us with a wealth of information regarding the distance, size, and orientation of surfaces in the environment.

3.3 Flow in the ambient optic array

> *What is clear to me now that was not clear before is that structure as such, frozen structure, is a myth, or at least a limiting case. Invariants of structure do not exist except in relation to variants.*
>
> *(Gibson, 1979, p.87)*

In the above quotation Gibson is highlighting the importance of another intrinsic aspect of perception that is often missing from laboratory stimuli – that of motion. His argument is that invariant information can only be perceived in relation to variant information, so to perceive invariant information we have to see the environment change over time.

There are two basic forms of movement: motion of the observer and motion of objects within the environment. Motion of the observer tends to produce the greatest degree of movement as the entire optic array is transformed (see Figure 3.9). Gibson suggested that this transformation provides valuable information about the position and shape of surfaces and objects. For example, information about shape and particularly position is revealed by a phenomenon known as **motion parallax**. The principle of motion parallax is that the further an object is from an observer, the less it will *appear* to move as the observer travels past it. Imagine the driver of a moving inter-city train looking out of their side-window at a herd of cows grazing in a large field next to the line. The cows near the train will *appear* to move past much faster than the cows at the back of the field. Thus, the degree of apparent motion is directly related to the distance of the object from the observer.

A second means by which observer motion can provide information about the shape and position of objects is through occlusion. Imagine the same observer described above travelling past the same field of cows. Their motion will cause the cows nearest to the train to pass in front of, or *occlude*, the cows grazing further away. This allows the observer to deduce that the *occluded* cows (i.e. the ones that become hidden by other cows) are further away than those doing the occluding.

Gibson dealt with the motion of the observer through reference to flow patterns in the optic array. As our train driver looks at the grazing cows by the side of the track, the entire optic array will appear to flow past from left to right, assuming that the driver looks out of the right-hand window (see Figure 3.14).

When the train driver becomes bored of cow watching and returns their attention to the track in front of the train, the flow patterns in the optic array will change so that the texture elements appear to be radiating from the direction in which the train is travelling (the apparent origin of this radiating flow pattern is known as the pole). The texture elements that make up the surfaces in the environment will appear to emerge from the pole, stream toward the observer, and then disappear from view (see Figure 3.15).

This pattern would be completely reversed if the guard at the rear of the train were to look back toward the direction from which the train had come (see Figure 3.16).

Gibson proposed a set of rules that linked flow in the optic array to the movement of the observer through the environment (Gibson, 1979):

- If there is flow in the ambient optic array, the observer is in motion; if there is no flow, the observer is not moving.
- Outflow of the optic array from the pole specifies approach by the observer and inflow to the pole specifies retreat.

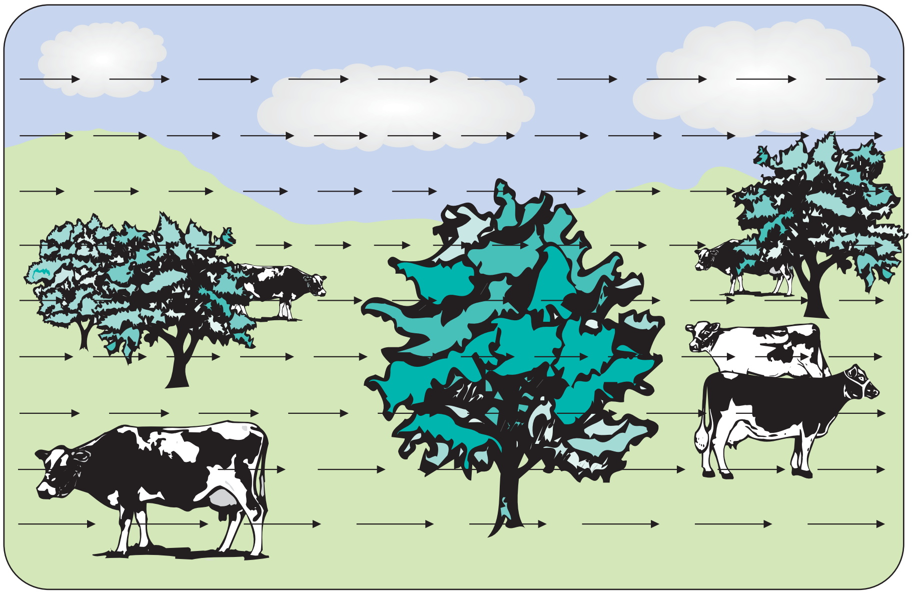


FIGURE 3.14 Flow patterns in the optic array parallel to the direction of the observer's motion.

- The direction of the pole specifies the direction in which the observer is moving.
- A change in the direction of the pole specifies that the observer is moving in a new direction.

For Gibson, the movement of the observer was a critical part of perception. In fact, he deemed it of such importance that he saw the *perceptual system* as not being limited to the eyes and other sense organs, but constituting a hierarchy of organs in which the eyes are linked to a head that can turn, which is linked to a body that can move. As Gibson said: 'Perceiving is an act, not a response, an act of attention, not a triggered impression, an achievement, not a reflex' (Gibson, 1979, p.21).

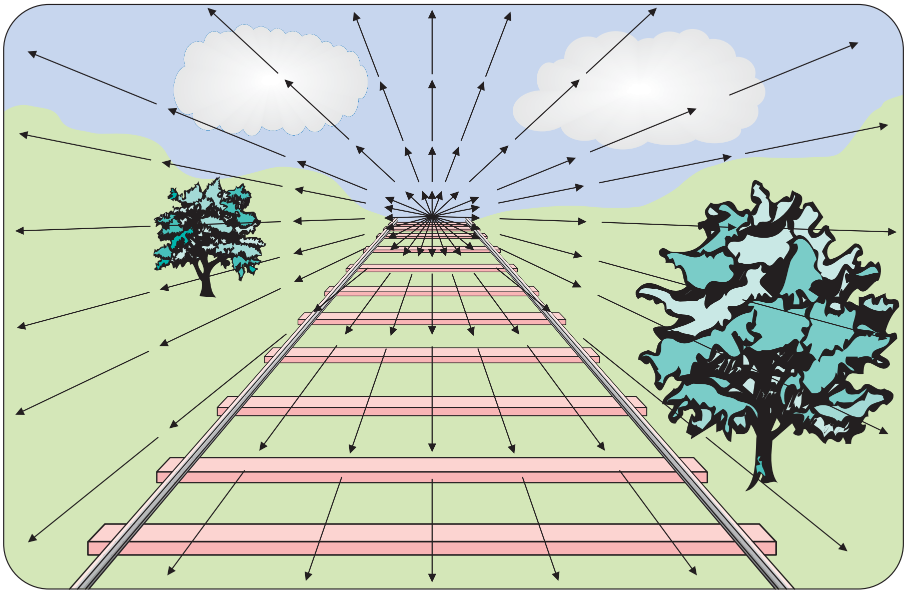


FIGURE 3.15 Flow patterns in the optic array in the direction of the observer's motion.

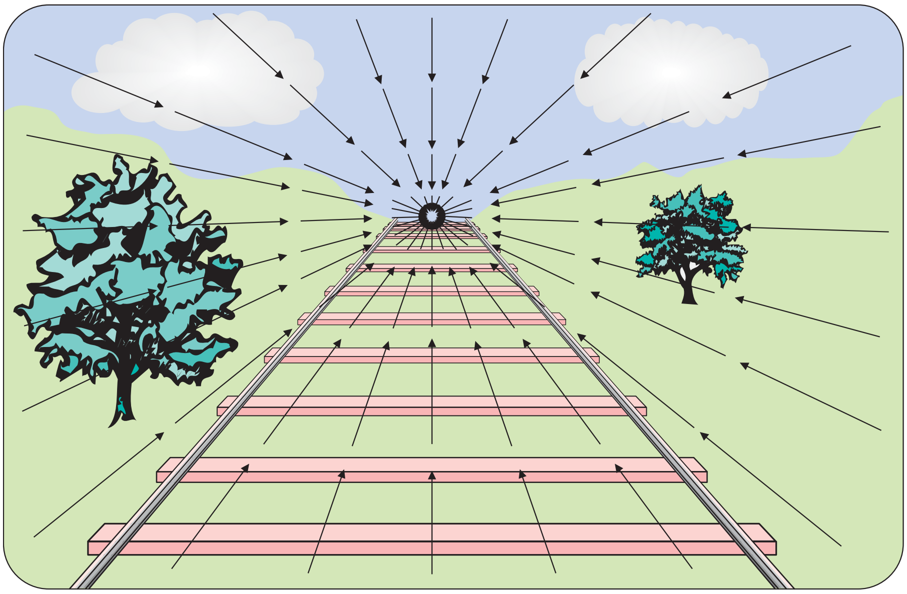


FIGURE 3.16 Flow patterns in the optic array in the opposite direction to the observer's motion.

3.4 Affordances and resonance

We began our discussion of Gibson's theory by stating that he saw information as being directly perceived or 'picked up' from the environment. In his later work Gibson (1979) took this idea of information being 'picked up' one step further and suggested that the end point of the perceptual process was not a visual description of the surrounding world, but rather that objects directly 'afforded' their use.

At its simplest (and least controversial) level, the concept of **affordance** builds on earlier research conducted by the Gestalt psychologists, in which the features of objects were seen as providing information as to their use. For instance, the features of a rock would suggest that it could be stood upon, the features of a fallen branch that it could be picked up, and the features of a fruit that it could be eaten.

However, Gibson makes two claims regarding affordances that are rather harder to accept and have proven to be far more controversial. First, he states that affordances act as a bridge between perception and action and do not require the intervention of any cognitive processes. Just as the nature of the environment can be directly 'picked up' from the structure of the optic array, the observer can interact with surfaces and objects in the environment directly through affordance.

Second, Gibson saw no role for memory in perception, as the observer does not have to consult their prior experience in order to be able to interact with the world around them. Instead he states that the perceptual system resonates to invariant information in the optic array. Although the definition of 'resonates' and the identity of what is doing the resonating is left very vague by Gibson, the point is that 'global' information about the optic array (in the form of invariant information) is dealt with by the perceptual system without the need to analyse more 'local' information such as lines and edges.

These assertions may seem unreasonable to you, as they have done to other researchers. If we are studying psychology, then surely the cognitive processes that allow us to perceive must be one focus of our attention. In addition, if when perceiving the world we do not make use of our prior experiences, how will we ever learn from our mistakes? In the next two sections we shall turn to theories that attempt to deal with these issues and to explain exactly how the brain makes sense of the world around us.

However, even if Gibson's theory does not enlighten us as to the nature of the cognitive processes that are involved in perception, it has still been extremely influential, and researchers in perception still need to bear in mind his criticisms of the laboratory approach that makes use of artificial stimuli:

> *Experiments using dynamic naturalistic stimuli can now be conducted, virtual scenes can be constructed, and images of brain activity while viewing these can be captured in a way that would have been difficult to envisage a century ago. However, the simulated lure of the screen (or even a pair of screens) should not blind experimenters and theorists to the differences that exist between the virtual and the real.*
>
> *(Wade and Bruce, 2001, p.105)*

SUMMARY OF SECTION 3

- Gibson developed an ecological approach to perception and placed great emphasis on the way in which real objects and surfaces structure light – he termed this the ambient optic array.
- He suggested that invariant information (such as texture gradient) could be 'picked up' from the optic array to provide cues as to the position, orientation, and shape of surfaces.
- Invariant information could also be revealed by motion, which produces variants such as flow patterns in the optic array.
- The importance of real surfaces and of motion led Gibson to suggest that perception could not be studied using artificial stimuli in a laboratory setting.
- Gibson did not see perception as a product of complex cognitive analysis, but suggested that objects could 'afford' their use directly.
- Interaction with the environment is at the heart of Gibson's theory; action is seen as the 'goal' of perception.

4 MARR'S THEORY OF PERCEPTION

> *... the detection of physical invariants, like image surfaces, is exactly and precisely an information-processing problem, in modern terminology. And second, he (Gibson) vastly underrated the sheer difficulty of such detection ... Detecting physical invariants is just as difficult as Gibson feared, but nevertheless we can do it. And the only way to understand how is to treat it as an information-processing problem.*
>
> *(Marr, 1982, p.30)*

As we stated previously, one criticism that has been levelled at Gibson's approach is that it does not explain in sufficient detail *how* information is picked up from the environment. To address this problem, a theory was needed that attempted to explain exactly how the brain was able to take the information sensed by the eyes and turn it into an accurate, internal representation of the surrounding world. Such a theory was proposed by David Marr (1982).

Before we look at Marr's theory, it is worth pointing out some of the similarities and differences between the approaches taken by Marr and Gibson. Like Gibson, Marr's theory suggests that the information from the senses is sufficient to allow perception to occur. However, unlike Gibson, Marr adopted an information-processing approach in which the processes responsible for analysing the retinal image were central. Marr's theory is therefore strongly 'bottom-up', in that it sees the retinal image as the starting point of perception and explores how this image might be analysed in order to produce a description of the environment. This meant that, unlike Gibson who saw action as the end point of perception, Marr concentrated on the perceptual processes involved in object recognition.

Marr saw the analysis of the retinal image as occurring in four distinct stages, with each stage taking the output of the previous one and performing a new set of analyses on it. The four stages were:

1. *Grey level description* – the intensity of light is measured at each point in the retinal image.
2. *Primal sketch* – first, in the raw primal sketch, areas that could potentially correspond to the edges and texture of objects are identified. Then, in the full primal sketch, these areas are used to generate a description of the outline of any objects in view.
3. *2½D sketch* – at this stage a description is formed of how the surfaces in view relate to one another and to the observer.
4. *3D object-centred description* – at this stage object descriptions are produced that allow the object to be recognized from any angle (i.e. independent of the viewpoint of the observer).

More generally, Marr concentrated his work at the computational theory and algorithmic levels of analysis (see Chapter 1) and had little to say about the neural hardware that might be involved. One reason for this is that he developed his theory largely by designing computer-based models and algorithms that could perform the requisite analyses.

4.1 The grey level description

One way of describing the first stage in Marr's theory is to say that it gets rid of colour information. This is not because Marr thought that colour was unimportant in perception. Rather, he thought that colour information was processed by a distinct **module** and need not be involved in obtaining descriptions of the shape of objects and the layout of the environment. In fact, the modular nature of perception was a fundamental part of Marr's theory:

> *Computer scientists call the separate pieces of a process its modules, and the idea that a large computation can be split up and implemented as a collection of parts that are as nearly independent of one another as the overall task allows, is so important that I was moved to elevate it to a principle; the principle of modular design.*
>
> *(Marr, 1982, p.102)*

This meant that the perception of colour could be handled by one 'module' and the perception of shape by another.

The first stage in Marr's theory acts to produce a description containing the intensity (i.e. the brightness) of light at all points of the retina. A description composed solely of intensity information is referred to as 'greyscale', because without the information provided by analysing the wavelength of light, it will consist of nothing but different tones of grey. If you turn down the colour on your TV, the resulting picture will be a greyscale image – although we call it 'black and white', it actually consists of many shades of grey.

Without going into too much detail, it is possible to derive the intensity of the light striking each part of the retina because as light strikes a cell in the retina, the voltage across the cell membrane changes and the size of this change (or **depolarization**) corresponds to the intensity of the light. Therefore, a greyscale (or grey level) description is produced by the pattern of depolarization on the retina. In other words, it is possible to derive the greyscale description simply by analysing the outputs of the receptor cells in the retina.

4.2 The primal sketch

The next part in Marr's theory, the generation of the primal sketch, occurs in two stages. The first stage consists of forming a raw primal sketch from the grey level description by identifying patterns of changing

✱ ACTIVITY 3.7

Find a wooden table or chair and place it where it is both well illuminated and against a light background. Describe how the intensity of the light reflected from the table/chair changes across its surface and in comparison with the background.

COMMENT

You should be able to see that the edges of the table/chair are marked by a quite large, sharp change in the intensity of the reflected light caused by the object in question being darker than the background. In addition, there are smaller changes in intensity caused by the individual parts of the table/chair and by the texture of the wood. You may also have noticed other changes in the intensity of the reflected light that did not correspond to the edge of the object, its parts, or its texture.

intensity. Activity 3.7 may help you understand what this means.

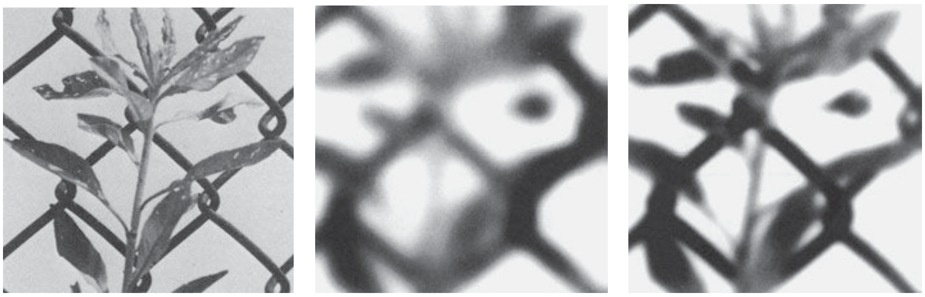
It is possible to group changes in the intensity of the reflected light into three categories:

- Relatively large changes in intensity produced by the edge of an object.
- Smaller changes in intensity caused by the parts and texture of an object.
- Still smaller changes in intensity due to random fluctuations in the light reflected.

Marr and Hildreth (1980) proposed an algorithm that could be used to determine which intensity changes corresponded to the edges of objects, meaning that changes in intensity due to random fluctuations could be discarded. The algorithm made use of a technique called **Gaussian blurring**, which involves averaging the intensity values in circular regions of the greyscale description. The values at the centre of the circle are weighted more than those at the edges in a way identical to a normal (or Gaussian) distribution.

By changing the size of the circle in which intensity values are averaged, it is possible to produce a range of images blurred to different degrees. Figure 3.17 shows images that have been produced in this manner. The original (i.e. unblurred) image is shown in (a). As you can see, using a wider circle (b) produces a more blurred image than using a narrower circle (c).

Marr and Hildreth's algorithm works by comparing images that have been blurred to different degrees. If an intensity change is visible at two or more adjacent levels of blurring, then it is assumed that it cannot correspond to a random fluctuation and must relate to the edge of an object. Although this algorithm was implemented by Marr and Hildreth on a computer, there is evidence that retinal processing delivers descriptions that have been blurred to different degrees.



(a) (b) (c)

FIGURE 3.17 Examples of Gaussian blurred images. *Source: Marr and Hildreth, 1980, p.190*

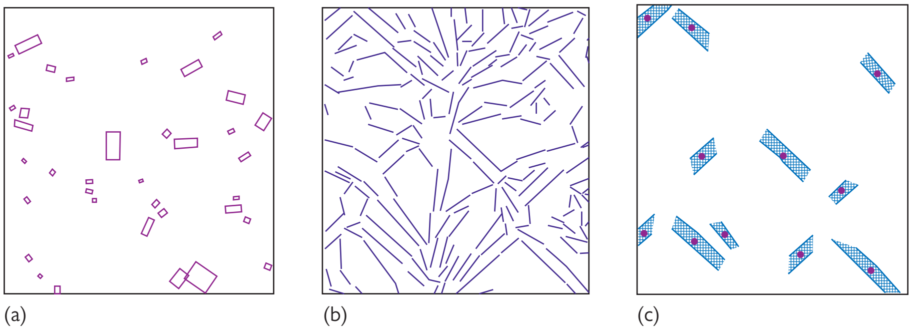


FIGURE 3.18 Primitives used in the raw primal sketch: (a) blobs; (b) edge-segments; and (c) bars. *Source: Marr, 1982, Figure 2.21, p.72*

By analysing the changes in intensity values in the blurred images, it is possible to form a symbolic representation consisting of four **primitives** corresponding to four types of intensity change. Marr referred to these primitives as 'edge-segments', 'bars', 'terminations', and 'blobs'. An edge-segment represented a sudden change in intensity; a bar represented two parallel edge-segments; a termination represented a sudden discontinuity; and a blob corresponded to a small, enclosed area bounded by changes in intensity. In Figure 3.18 you can see how the image shown in Figure 3.17(a) would be represented using three of these primitives, whilst Figure 3.19 shows how three simple lines would be represented in the raw primal sketch.

As you can see from Figure 3.19, although the raw primal sketch contains a lot of information about details in the image, it does not contain explicit information

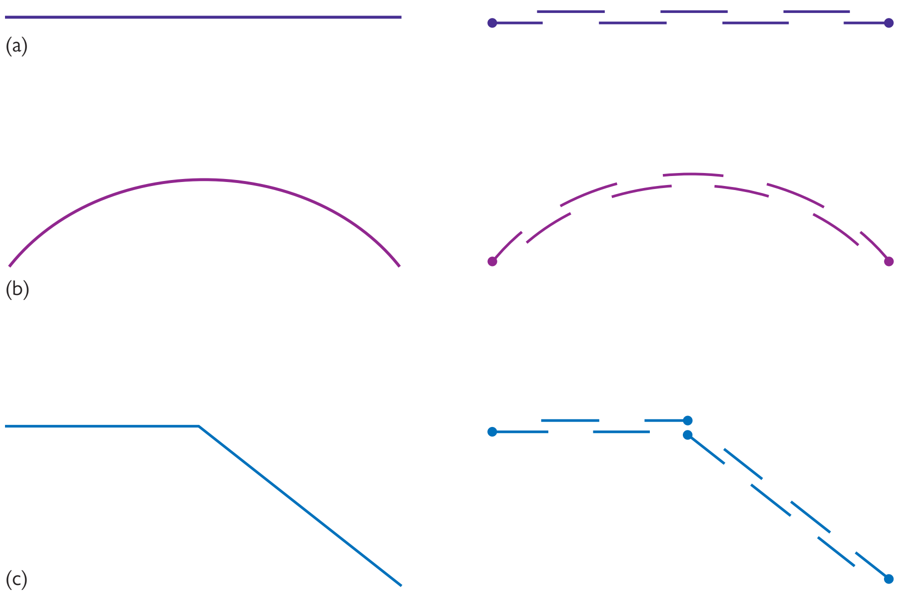


FIGURE 3.19 Representation of three simple lines in the raw primal sketch: 'The raw primal sketch represents a straight line as a termination, several oriented segments, and a second termination (a). If the line is replaced by a smooth curve, the orientations of the inner segments will gradually change (b). If the line changes its orientation suddenly in the middle (c), its representation will include an explicit pointer to this discontinuity. Thus in this representation, smoothness and continuity are assumed to hold unless explicitly negated by an assertion'. *Source: Marr, 1982, Figure 2.22, p.74*

about the global structure of the objects in view. The next step is therefore to transform the raw primal sketch into a description, known as the **full primal sketch**, which contains information about how the image is organized, particularly the location, shape, texture, and internal parts of any objects that are in view.

Basically, the idea is that place tokens are assigned to areas of the raw primal sketch based on the grouping of the edge-segments, bars, terminations, and blobs. If these place tokens then form a group themselves, they can be aggregated together to form a new, higher-order place token.

Imagine looking at a tiger. The raw primal sketch would contain information about the edge of the tiger's body, but also about the edges and pattern of its stripes and the texture of its hair. In the full primal sketch, place tokens will be produced by the grouping of the individual hairs into each of the stripes. The place tokens for each stripe would then also be grouped (because they run in a consistent vertical pattern along the tiger) into a higher-order place token, meaning that there will be at least two levels of place tokens making up the tiger.

Various mechanisms exist for grouping the raw primal sketch components into place tokens and for grouping place tokens together. These include clustering, in which tokens that are close to one another are grouped in a way very similar to the Gestalt principle of proximity, and curvilinear aggregation, in which tokens with related alignments are grouped in a similar fashion to the Gestalt principle of good continuation.

As we saw in Section 2, perceptual grouping is a robust, long-established, and powerful effect. Marr saw algorithms expressing laws such as those formulated by the Gestalt approach as being responsible for turning the ambiguous raw primal sketch into the full primal sketch in which the organization of objects and surfaces was specified.

4.3 The 2½D sketch

In Marr's theory, the goal of early visual processing is the production of a description of the environment in which the layout of surfaces and objects is specified in relation to the particular view that the observer has at that time. Up until now we have been looking at how the shape of objects and surfaces can be recovered from the retinal image. However, in order to specify the layout of surfaces, we need to now include other information, specifically cues that tell us how far away each surface is.

Marr's modular approach to perception means that while the full primal sketch is being produced, other visual information is being analysed simultaneously. Much of this has to do with establishing depth relations, the distance between a surface and the observer, and also how far objects extend. We saw in Section 3 that motion cues and cues from texture can be used to specify the distance to an object, and it is also possible to make use of the disparity in the retinal images of the two eyes (known as **stereopsis**), and shading cues that are represented in the primal sketch.

Marr proposed that the information from all these 'modules' was combined together to produce the **2½D sketch**. It is called the 2½D sketch rather than the 3D sketch because the specification of the position and depth of surfaces and objects is done in relation to the observer. Thus, the description of an object will be **viewer-centred** and will not contain any information about the object that is not present in the retinal image. How the viewer-centred 2½D sketch is turned into a fully 3D, **object-centred** description is one of the topics dealt with in the next chapter.

Marr saw the 2½D sketch as consisting of a series of primitives that contained vectors (a line depicting both size and direction) showing the orientation of each surface. A vector can be seen as a needle, in which the direction the needle is pointing tells us in which direction the surface is facing, and its length tells us by how much the surface is slanted in relation to the observer. A cube would therefore be represented like the one shown in Figure 3.20. In addition to the information shown in Figure 3.20, Marr suggested that each vector (or needle) would have a number associated with it that indicated the distance from the observer.

The 2½D sketch therefore provides an unambiguous description of the size, shape, location, orientation,

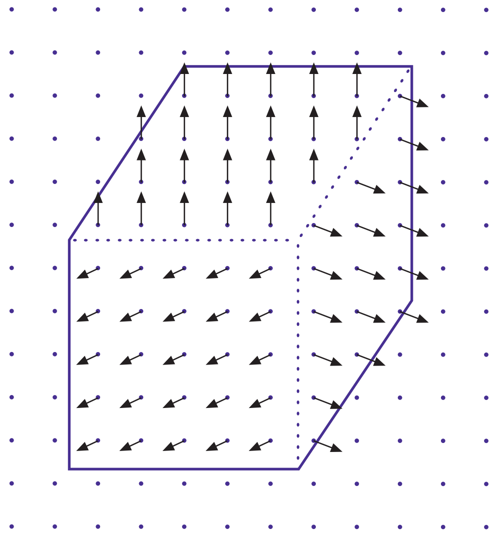


FIGURE 3.20 A $2\frac{1}{2}$D sketch of a cube. *Source: Marr, 1982, Figure 4.2, p.278*

and distance of all the surfaces currently in view, in relation to the observer.

4.4 Evaluating Marr's approach

Marr's theory was the catalyst for a great deal of computational and psychological research. Some of this research has reported findings consistent with the mechanisms proposed by Marr, whilst some has found that Marr's theory does not offer a good explanation for the results obtained. We will not attempt to review every single study here, but instead describe a few studies that have tested elements of Marr's theory.

Marr and Hildreth (1980) attempted to test their idea that the raw primal sketch was formed by searching for changes in intensity values in adjacent levels of blurring, by implementing this algorithm in a computer program. They found that when applied to images of everyday scenes the algorithm was reasonably successful in locating the edges of objects. However, as with all computer-simulation research, it is important to remember that, just because a specific program yields the expected results, it does not necessarily follow that this is what is happening in the human perceptual system.

It seems as if Marr's approach to the formation of the full primal sketch was flawed in that it was limited to grouping strategies based on the 2D properties of an image. Enns and Rensick (1990) showed that participants could easily determine which one of a series of figures consisting of blocks was the odd one out, even though the only difference between the figures was their orientation in three dimensions. Thus, some grouping strategies must make use of 3D information.

One area in which Marr's theory does seem to fit the results of experimentation is in the integration of depth cues in the $2\frac{1}{2}$D sketch, studied in experiments that have attempted to isolate certain forms of depth cue and then determine how they interact. For example, Young *et al.* (1993) looked at how motion cues interacted with texture cues. They concluded that the perceptual system does process these cues separately, and will also make selective use of them depending on how 'noisy' they are. In other words, in forming the $2\frac{1}{2}$D sketch, the perceptual system does seem to integrate different modules of depth information, but will also place more emphasis on those modules that are particularly useful for processing the current image.

As well as the success of the specific processes suggested by Marr, it is also possible to evaluate his theory according to broader concepts. As we shall see in Section 6, there is evidence that there are two visual pathways in the brain that appear to process separately 'what' information and 'where' information. It seems that different perceptual processes exist according to whether the goal of perception is action or object recognition. Although Marr's theory is a modular approach, so that different types of visual information are processed separately, it did not predict the separation of visual pathways into action and object recognition, and indeed it is hard to incorporate this into the theory (Wade and Bruce, 2001). However, although the precise nature of the processes suggested by Marr may not map exactly onto those actually used by the brain to perceive the world, the impact of Marr's theory should not be underestimated: 'Thus it is not the details of Marr's theory which have so far stood the test of time, but the approach itself' (Wade and Bruce, 2001, p.97).

SUMMARY OF SECTION 4

- Marr proposed a theory of vision that was based on bottom-up processing of information.
- His approach was to see perception as being composed of a series of stages, with each stage generating an increasingly sophisticated description.
- Marr saw the end point of the perceptual process as object recognition rather than action.
- The first stage involves producing a grey level description based on the activation of retinal cells.
- This description is analysed by blurring it to different degrees. Changes in intensity value that are present in two or more adjacent levels of blurring are assumed to correspond to the 'edge' of an object (or part of an object).
- The raw primal sketch is generated by assigning one of four primitives (edge-segment, bar, termination, or blob) to each change in intensity values.
- The full primal sketch is generated by using perceptual organizational principles such as clustering and similarity to group these primitives together and assign each group a place token.
- Information from different modules (such as stereopsis and motion) are combined with the full primal sketch to produce the $2\frac{1}{2}$D sketch. This contains primitives consisting of vectors that reveal the distance and orientation, in relation to the observer, of the visible surfaces.

5 CONSTRUCTIVIST APPROACHES TO PERCEPTION

The previous sections of this chapter should have given you some idea of how we can see and interpret sensory information. The emphasis so far has been on 'bottom-up' processes (see also Activity 3.8). As discussed previously, there is also information flowing 'top-down' from stored knowledge. This makes intuitive sense. To be able to perceive something as 'a bus', you need to access stored knowledge concerning what the features of a bus actually are (big object with wheels, etc.).

Thus, what you see a stimulus *as* depends on what you know. This notion, that perceiving something involves using stored knowledge as well as information coming in from the senses, is embodied in an approach referred to as the **constructivist approach**. The approach is described as 'constructivist' because it is based on the idea that the sensory information that forms the basis of perception is, as we have already suggested, incomplete. It is necessary to *build* (or 'construct') our perception of the world from *incomplete* information. To do this we use what we already *know* about the world to interpret the incomplete sensory information coming in and 'make sense' of it. Thus stored knowledge is used to aid in the recognition of objects.

Two of the foremost proponents of the constructivist approach are Irvin Rock (1977, 1983, 1997) and Richard Gregory (1980). Gregory suggested that individuals attempt to recognize objects by generating a series of **perceptual hypotheses** about what that object might be. Gregory conceptualized this process as being akin to how a scientist might investigate a problem by generating a series of hypotheses and accepting the one that is best supported by the data (in perception, 'data' would be the information flowing 'up' from the senses).

✱ ACTIVITY 3.8

Look back at Activity 3.1. Can you explain any of the visual illusions in terms of what you now know about the bottom-up approach to perception?

COMMENT

Gibson would tell us that the Necker cube is a geometric figure that contains none of the information (particularly texture gradients) that we would usually use when perceiving an object. Marr's theory can help us to explain Kanizsa's illusory square, as the four areas of intensity change corresponding to the missing parts of the circles would be grouped together to form a square.

But what about the Müller–Lyer illusion? There are a number of alternative explanations for this illusion, one of which is that we group each vertical line with its set of arrowheads to form a single object. This of course results in the object with the inward-pointing arrowheads being larger than the one with the outward-pointing arrowheads; basically, due to perceptual grouping we cannot separate the vertical line from the overall size of the object. However, as the Müller–Lyer illusion is reduced if the straight arrowheads are replaced with curved lines (see Figure 3.21), it could be that we also need to look at an explanation based on top-down perception.

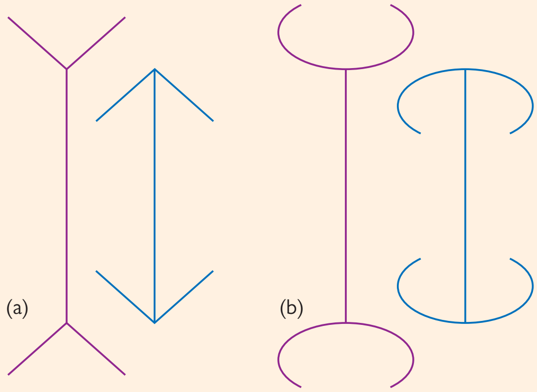


FIGURE 3.21 The original Müller-Lyer illusion (a), and with curved arrowheads (b).

As we saw in Activity 3.1, another explanation of the Müller–Lyer illusion is that we make use of top-down information and see the outward-pointing arrowheads as an indication that the vertical line is nearer to us than the line with the inward-pointing arrowheads.

We are forced to generate hypotheses, according to Gregory's argument, because the sensory data are incomplete. If we had perfect and comprehensive sensory data we would have no need of hypotheses as we would *know* what we perceived. Stored knowledge is assumed to be central to the generation of perceptual hypotheses as it allows us to fill in the gaps in our sensory input. The influence of stored knowledge in guiding perceptual hypotheses can be demonstrated by the use of impoverished figures such as the one in Figure 3.22 (Street, 1931).

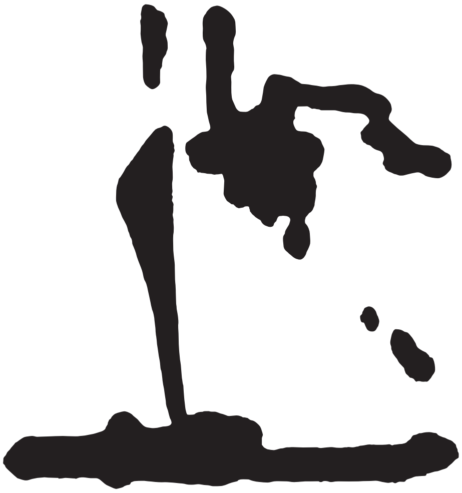


FIGURE 3.22 An example of an impoverished figure.
Source: Street, 1931

At first glance this picture may be difficult to perceive as anything other than a series of blobs. So the resulting hypothesis might be that it is just 'a load of blobs.' If, however, you are told that it is a picture of an ocean liner (coming towards you, viewed from water

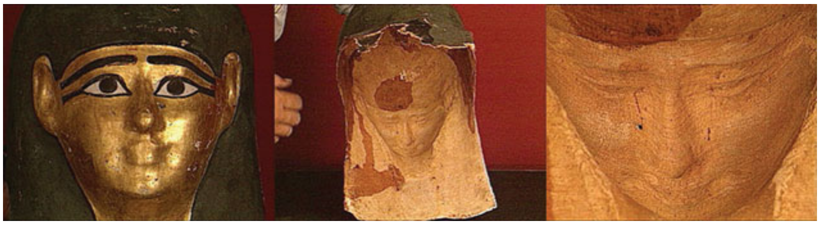


FIGURE 3.23 The mask of Hor.

level) then the picture may immediately resolve into an image of an ocean liner. The sensory information has not changed, but what you know about it has, allowing you to generate a reasonable hypothesis of what the figure represents. Similarly, in the example used in Activity 3.3 of trying to identify an object by touch alone, if you are given some clues about the function of the object (i.e. your knowledge related to the object is increased), it is likely to be easier to identify it.

The use of knowledge to guide our perceptual hypotheses may not always lead to a 'correct' perception. There are some stimuli with which we are so familiar (such as faces) that there can be a strong bias towards accepting a particular perceptual hypothesis, resulting in a 'false' perception. For instance, look at the faces in Figure 3.23.

This is the mask of Hor, an Egyptian mummy. The first view is the mask from the front and the second two are of the back of the mask. Although the face viewed from the back is 'hollow' it still appears perceptually as a normal face. Our knowledge of how a face is supposed to look is (according to Gregory, 1980) so strong that we cannot accept the hypothesis that a face could be 'hollow.' This effect is interesting in that it provides an example of a perceptual hypothesis conflicting with what Gregory terms 'high-level' knowledge. You *know* at a conceptual level that the mask is hollow, yet you still *perceive* it as a 'normal' face. This, as Gregory suggests, represents a tendency to go with the most *likely* hypothesis. The Penrose triangle (Penrose and Penrose, 1958) in Figure 3.24 demonstrates a similar point.

It would be impossible to construct the object in Figure 3.24 so that the three sides of the triangle were joined. At one level, we 'know' that this must be true. Yet whichever corner of the triangle we attend to suggests a particular 3D interpretation. Our interpretation of the figure changes as our eyes (or just our attention) jumps from corner to corner. These data-supported interpretations, or hypotheses, tend to overwhelm the conceptual knowledge that we are viewing a flat pattern.

Although the constructivist approach in general, and Gregory's theories in particular, provide an attractive explanatory framework for perception, there are areas of the theory (as there were with Gibson's approach) that are left rather vague. For instance, how do we actually generate hypotheses and how do we know when to stop and decide which is the 'right' one? Why does knowledge sometimes but not always help perception? How can we 'know' something is wrong and yet still perceive it as wrong (as with the hollow face)? Although these are difficult questions to answer, progress is being made in explaining how human perception may be based, at least in part, on constructivist principles; some of this work will be discussed below.

Thus, there appears to be evidence that perceptions of the outside world can be 'constructed' using information flowing 'up' from the senses combined with knowledge flowing 'down'. However, this seems to be in direct contrast to the theories of Gibson and Marr

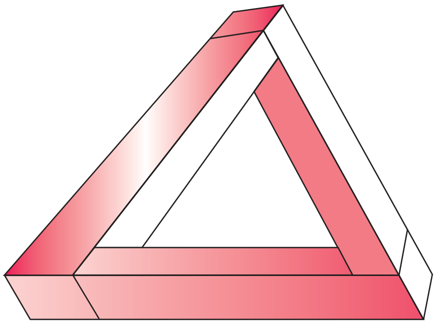


FIGURE 3.24 An impossible triangle. *Source: Penrose and Penrose, 1958*

discussed earlier, which suggest that there is no need to use stored knowledge to interpret the information flowing in from the senses. Indeed, the impossible triangle above shows that we do not always make use of knowledge that may be relevant and available. So, just how important is knowledge to the process of perception, and is there any way in which we can reconcile theories of perception that see knowledge as being essential with those that see it as unnecessary? The following section considers how these different theories may be reconciled through consideration of the way in which the brain processes sensory information.

SUMMARY OF SECTION 5

- What you see a stimulus as depends on what you know. This means that perception must involve top-down processing.
- The constructivist approach to perception is based on the idea that sensory data is often incomplete, so a description can only be constructed by making use of stored knowledge.
- Gregory suggested that sensory data are incomplete and we perceive by generating a series of perceptual hypotheses about what an object might be.
- The use of stored information can lead to perceptual hypotheses that are inaccurate, which is why we may be fooled by some visual illusions.

6 THE PHYSIOLOGY OF THE HUMAN VISUAL SYSTEM

There appear to be at least two (and maybe more) partially distinct streams of information flowing back from the retina (via the optic nerve) into the brain (e.g. Shapley, 1995). The characteristics of these streams and their relation to the theories of perception already described is the topic of this section. It should be emphasized that the distinction between the two streams is fairly loose. There is overlap in the types of information that the streams carry and there are numerous interconnections between them, but they may conveniently be conceptualized as distinct. The following subsections trace these streams of information from the retina to the brain.

6.1 From the eye to the brain

You may remember from Section 1.2 that there are two types of light-sensitive cells in the retina, called rods and cones. Both rods and cones are connected to what are termed retinal ganglion cells that essentially connect the retina to the brain. Ganglion cell axons leave the eye via the 'blind spot' (the concentration of blood vessels and nerve axons here means that there is no room for any receptors, hence this region is 'blind'). These cells then project (send connections) to an area termed the lateral geniculate nucleus (LGN), and from there to the area of the brain known as the 'primary visual cortex' (also known as V1). Even at the level of retinal ganglion cells, there is evidence of two distinct streams or 'pathways', referred to as the parvocellular pathway and the magnocellular pathway (e.g. Shapley, 1995). These names derive from the relative sizes of the cells in the two pathways, with larger cells in the magnocellular pathway and smaller cells in the parvocellular one. This distinction is maintained up to and within the primary visual cortex, although there are interconnections between the two pathways.

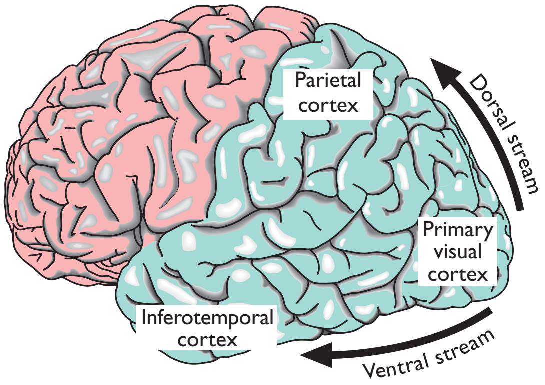


FIGURE 3.25 The dorsal and ventral streams.

Information travelling onward from the primary visual cortex is still maintained in two distinct streams (see Figure 3.25). One stream, leading to the inferotemporal cortex, is termed the **ventral stream**, and the other, leading to the parietal cortex, is known as the **dorsal stream** (these were described briefly in Chapter 2, Section 5.1).

6.2 The dorsal and ventral streams

The *ventral* stream projects to regions of the brain that appear to be involved in pattern discrimination and object recognition, whilst the *dorsal* stream projects to areas of the brain that appear to be specialized for the analysis of information about the position and movement of objects. Schneider (1967, 1969) carried out work with hamsters that suggested that there were two distinct parts of the visual system, one system concerned with making pattern discriminations and the other involved with orientation in space. Schneider suggested that one system is concerned with the question, 'What is it?', whereas the other system is concerned with the question, 'Where is it?'. This, and later work (Ungerleider and Mishkin, 1982), led to the ventral pathway being labelled a 'what' system, and the dorsal pathway a 'where' system.

Although the two streams appear to be specialized for processing different kinds of information, there is ample evidence of a huge degree of interconnection between the systems at all levels. Also, the streams appear to converge in the prefrontal cortex (Rao *et al.*, 1997), although there is still some evidence that the dorsal–ventral distinction is maintained (Courtney *et al.*, 1996). It has been suggested that it is in the prefrontal cortex that meaning is associated with the information carried by the two streams.

Although describing the two streams as 'what' and 'where' is convenient, there is a large body of work that suggests that the distinction is not quite that straightforward. For instance, Milner and Goodale (1995) report a number of studies with a patient, DF, who suffered severe carbon monoxide poisoning that appeared to prevent her using her ventral system for analysing sensory input. She could not recognize faces or objects, or even make simple visual discriminations such as between a triangle and a circle. She could draw objects from memory but not recognize them once she had drawn them. DF did, however, appear to have an intact dorsal stream. Although unable to tell if two discs were of the same or different widths (or even *indicate* the widths by adjusting the distance between her fingers), if she was asked to pick the discs up then the distance between her index finger and thumb as she went to pick them up was highly correlated with the width of the discs. In other words, she did *not* have size information available to conscious perception (via the ventral stream), but it *was* available to guide *action* (via the dorsal stream).

Norman (2002), following on from similar suggestions by Bridgeman (1992) and Neisser (1994), has drawn on the ongoing debate concerning the characteristics of the dorsal and ventral streams and suggested a dual-process approach. In this approach, the two streams are seen as acting synergistically so that the dorsal stream is largely concerned with perception for action and the ventral stream essentially concerned with perception for recognition. The dual-process approach is supported by some of the characteristics of the two streams (Norman, 2001, 2002):

1. There appears to be evidence (Goodale and Milner, 1992; Ungerleider and Mishkin, 1982) to suggest that the ventral stream is primarily concerned with recognition whilst the dorsal

stream drives visually-guided behaviour (pointing, grasping, etc.).

2. The ventral system is generally better at processing fine detail (Baizer *et al.*, 1991) whereas the dorsal system is better at processing motion (Logothesis, 1994).
3. The studies on patient DF (Milner and Goodale, 1995) suggest that the ventral system is knowledge-based and uses stored representations to recognize objects, whilst the dorsal system appears to have only very short-term storage available (Bridgeman *et al.*, 1997; Creem and Proffitt, 1998).
4. The dorsal system receives information faster than the ventral system (Bullier and Nowak, 1995).
5. A limited amount of psychophysical evidence suggests that we are much more conscious of ventral than of dorsal stream functioning (Ho, 1998).
6. It has been suggested (Goodale and Milner, 1992; Milner and Goodale, 1995) that the ventral system recognizes objects and is thus object-centred. The dorsal system is presumed to be used more in driving some action in relation to an object and thus uses a viewer-centred frame of reference (this distinction arises again in the next chapter).

6.3 The relationship between visual pathways and theories of perception

We have already seen that Gibson's approach to perception concentrated more on perception for the purposes of action, whilst Marr's theory was principally concerned with object recognition. The constructivist approach is also more concerned with perception for recognition than perception for action, as it concentrates on how we may use existing knowledge to work out what an object might be. Although these approaches have their differences, it is undoubtedly the case that we need to both recognize objects and perform actions in order to interact with the environment. It could be, then, that the type of perception discussed by Gibson is principally subserved by the dorsal system, whilst the ventral system is the basis for the recognition approach favoured by Marr and the constructivists.

For example, Gibson's notion of 'affordance' emphasizes that we might need to detect what things are *for* rather than what they actually *are*. That is, affordances are linked to actions ('lifting' or 'eating', for example). The dorsal system appears to be ideally suited to providing the sort of information we need to act in the environment. In addition, if a system is to be used to drive action, it really needs to be fast, as the dorsal stream seems to be.

The earlier discussion of Gibson's ecological approach also stated that Gibson saw no need for memory in perception. Certainly, one of the characteristics of the dorsal stream is that it appears to have no more than a very short 'memory' (at least for representations of objects). Thus, there appear to be some grounds for suggesting that the dorsal stream is Gibsonian in operation.

In contrast, the ventral stream appears to be ideally suited to the role of recognizing objects. It is specialized in analysing the sort of fine detail that Marr saw as essential to discriminating between objects, and it also seems able to draw on our existing knowledge (top-down information) to assist in identifying them. In addition, it is slower than the dorsal stream, but then recognizing what an object may be is not necessarily an immediate priority. For example, knowing that an object is *moving* towards you quickly is initially more important than knowing *what* it is.

6.4 A dual-process approach?

Norman's proposal discussed above does provide an attractive way of reconciling two of the classic approaches to visual perception. There is perhaps a danger, however, in trying to 'shoehorn' what is known about the dorsal and ventral streams into the framework provided by previous theories. Given that both the constructivist and Gibsonian theories are rather vague on how the processes they describe could be implemented, it is questionable how useful they are as a theoretical framework in which to interpret the workings of the dorsal and ventral streams. Attempting to explain the streams in the light of the previous theories does tend to emphasize the way in which they work separately rather than the way in which they work together. Undoubtedly, the two streams *can* operate independently (as demonstrated by the case of DF discussed earlier), but this is rather like saying that you can take the steering column out of a car and both the car and the steering wheel will still function to some degree! In fact, Norman (2002) describes the two streams as synergistic and interconnected, rather than independent.

Binsted and Carlton (2002), in a commentary on the proposal put forward by Norman, provide an illustration of the interaction between the dorsal and ventral streams using the example of skill acquisition. Previous work (Fitts, 1964) suggests that the early stages of learning a skill (such as driving) are characterized by cognitive processes of the sort associated with the ventral stream, whereas once the task is well practised it is characterized by learned motor actions of the sort associated with the dorsal stream.

The question is, if these two streams function in such different ways, how is learning transferred from one to the other? It is possible, of course, that learning occurs in both streams at the same time and that whichever is most effective 'leads' in performance of the task, but this still implies a high degree of interaction between them and a blurring of the boundaries between their functions. The issue (which is as yet unresolved) then becomes whether the two streams interact to such an extent that it is meaningless to consider them to be functionally separate and representative of different theoretical approaches to visual processing (as Norman suggests). Thus, rather than questioning whether both Gibsonian and constructivist principles are operating in visual processing, the debate centres on whether it is appropriate to ascribe these types of processing to discrete pathways. Whatever the outcome of the debate, Norman does present a compelling argument that visual processing does *not* have to be either for action *or* for recognition; it can be both.

6.5 Applying perceptual research: a case study of an 'aircraft proximity event'

One of the themes that runs throughout this chapter is the importance of motion in visual perception. In Section 3.3, two basic types of motion are discussed: motion of the observer and motion of objects in the environment. An indication of how we detect change in our environment comes from studies such as that of Beck *et al.* (2005) where repetitive transcranial magnetic stimulation was used to disrupt activity in the right parietal cortex (a part of the dorsal stream). They found that when the activity of the dorsal stream was disrupted, there was also a disruption in the ability to detect changes in a visual stimulus.

One of the (many) roles the dorsal stream has developed could, therefore, be to detect changes in the environment generated by the motion of approaching objects (such as sabre tooth tigers or oncoming cars). Once an oncoming object is detected, it is then necessary for the observer to assess what it is and whether it may be a threat, and this is where the ventral stream comes into play.

A bit of introspection will also tell us that we are inherently very good at detecting sudden changes in our environment, such as when something suddenly moves (or if there is a sudden new noise). Think of how many times you have suddenly become aware of something moving 'out of the corner of your eye' (it's usually a spider!). One thing that we might deduce from this is that if something could creep up on us without appearing to move, then it should be much more difficult to detect. This sounds impossible, but can happen if the motion of the observer and the motion of the object effectively cancel each other out. This is precisely how some air accidents occur, when pilots are relying on visual detection of approaching aircraft.

If two aircraft are flying on converging flight paths at constant speeds, as shown in Figure 3.26, the bearing of the aircraft relative to each other (i.e. the angle between them) will remain constant. This means that there appears to be no relative motion of one aircraft

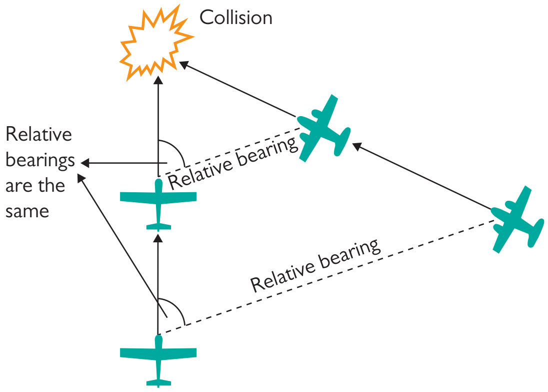


FIGURE 3.26 Two aircraft approaching on straight paths, at a constant speed, and on a collision course. Each plane would not appear to be moving relative to the other, making detection less likely.

with respect to the other. If the pilots are relying on visual 'see and avoid' principles, then an increase in perceived size of the other aircraft will be the only clue that it is getting closer. This lack of relative motion is recognized as a contributory factor in some air accidents and near misses. An example of this was an 'aircraft proximity event' (aka 'near miss') at Bankstown Aerodrome in New South Wales (ATSB, 2009). In this example, two aircraft (a Chieftan and a Cherokee) were approaching the airport on converging tracks and neither appeared aware of the other until the Chieftan pilot took 'severe avoiding action' at the last moment and narrowly managed to avoid a collision. The pilot of the Cherokee also became aware of the other aircraft, but only 1–2 seconds before the possible collision, and was unable to react to the situation. Although, theoretically, these two aircraft should have been able to see each other, the lack of relative motion (as indicated in Figure 3.26) made it much more difficult for the pilots to perceive the potential collision hazard. The investigation into the incident cited as a key finding that, 'Both aircraft would have appeared to be stationary objects in each pilot's visual field, decreasing the likelihood that the respective pilots would sight the other aircraft' (ATSB, 2009, p.6).

This aircraft proximity event also illustrates another aspect of perception embodied by the constructivist approach; that registering sensory information does not guarantee that it will make sense. The sensory information may be registered perfectly, but the resulting hypothesis as to what it means may be wrong. In the above incident, Air traffic control (ATC) attempted to advise the Cherokee to take avoiding action using the phrase, '…widen out to the left …'. Obviously the air traffic controller knew what was meant by that instruction (the Cherokee was being instructed to turn to the left in a wider circle). This was not, however, understood by the (relatively inexperienced) pilot of the Cherokee, who failed to take the avoiding action indicated by the ATC.

So, in 'real life' just as in the lab, perception is not always perfect. Luckily in this case there was no collision, although the pilot of the Chieftain did report that '… the sudden evasive manoeuvre caused the pilot's head to hit the roof of the cockpit…'!

6.6 Combining bottom-up and top-down processing

As we have shown, approaches to perception can be differentiated according to whether they are primarily concerned with perception for action or recognition, or with bottom-up or top-down processing. It may have occurred to you when reading about these approaches that it is likely that perception must in fact contain elements of both types of processing. A key question, then, is whether there is any evidence that this is in fact the case.

You were introduced to the idea of visual masking in the last chapter, particularly the concept of backward masking, in which the presentation of a second image disrupted the perception of an initial image. In Figure 3.27 you can see sets of stimuli that have been used to demonstrate two different types of visual masking. In each case, the mask is presented after a very brief presentation of the target. The task facing the participant is to report which corner of the diamond target is missing.

Standard explanations of why masking occurs with the stimuli in Figure 3.27 require that the mask contains contours that either overlap (Figure 3.27(a)) or exactly coincide with (Figure 3.27(b)) those of the

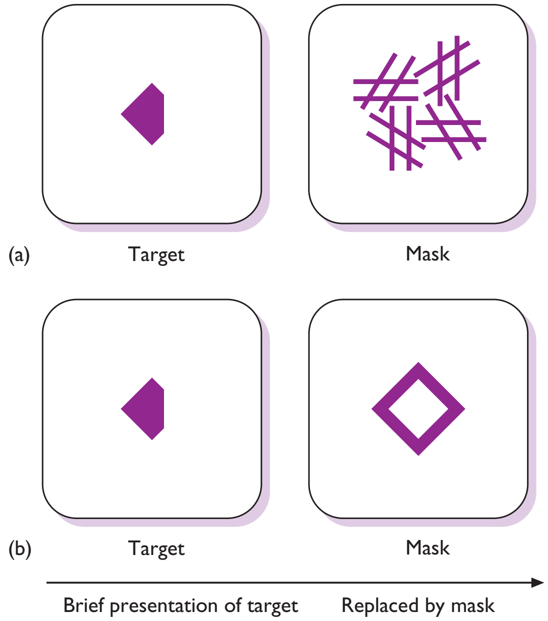


FIGURE 3.27 Stimuli used to demonstrate backward masking.

target (Enns and Di Lollo, 2000). But if masking is a product of the close similarity between the contours of target and mask, it is hard to account for the fact that a masking effect is also found for the images in Figure 3.28 (Di Lollo *et al.*, 1993).

Enns and Di Lollo (1997) reported that the four-dot pattern shown in Figure 3.28 appeared to mask the target if target and mask were presented together and the target displayed very briefly, or if the mask was displayed very soon after a brief presentation of the target. Enns and Di Lollo (2000) explained the masking observed using the four-dot pattern by reference to re-entrant processing. We know from neuroscience research that communication between two different regions of the brain is never unidirectional. If one region is sending a signal to another, then the second region also sends a signal back through what are referred to as re-entrant pathways (Felleman and Van Essen, 1991).

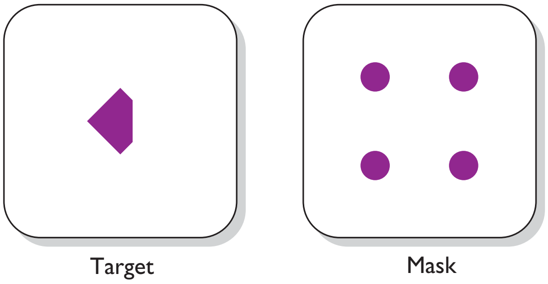


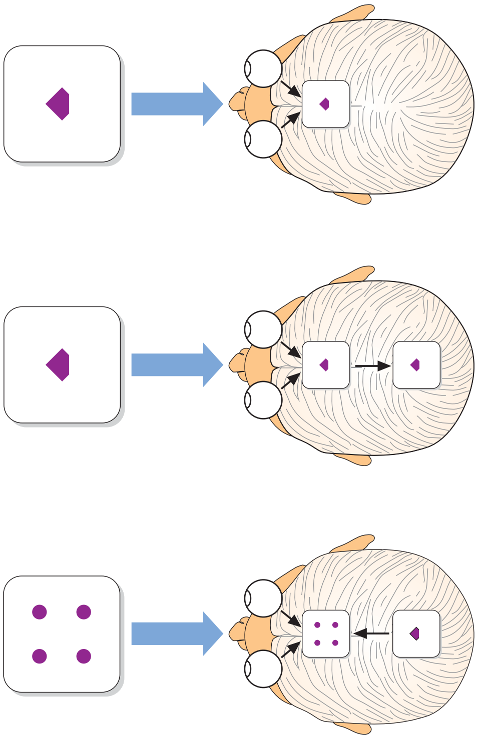
FIGURE 3.28 An example of a four-dot mask.

Hupe *et al.* (1998) suggested that re-entrant pathways could be used to allow the brain to check a perceptual hypothesis against the information in an incoming signal. In other words:

- Bottom-up processing produces a low-level description.
- This is used to generate a perceptual hypothesis at a higher level.
- Using re-entrant pathways, the accuracy of the perceptual hypothesis is assessed by comparing it with the (perhaps now changed) low-level description.

Di Lollo *et al.* (2000) used this idea as the basis for an explanation of visual masking. The idea is that each part of the displayed image(s) is perceived in terms of a combination of high-level descriptions similar to a perceptual hypothesis, and low-level codes produced by bottom-up processes. If the target is only presented very briefly, then masking can occur because by the time the high-level perceptual hypothesis is compared with the low-level bottom-up description, the target will have been replaced by the mask. Thus, the perceptual hypothesis will be rejected because it is based on a pattern (the target) that is different from the pattern currently being subjected to bottom-up processing (the mask) – see Figure 3.29.

The re-entrant processing explanation of visual masking is based upon the presumed interaction of bottom-up processes with top-down processes. This is consistent with the idea that perception is neither entirely bottom-up nor entirely top-down, but is actually reliant on both forms of processing. However, as the visual system is very complicated, it is likely that as well as re-entrant processing there are other processes involving top-down and bottom-up interactions contributing to four-dot masking (Gellatly *et al.*, 2010; Pilling and Gellatly, 2010).



Participant forms low-level description of target using bottom-up processing

Participant forms the perceptual hypothesis 'The image is of a diamond'

Participant checks the perceptual hypothesis against the current low-level description – but this is now of the four circles. The hypothesis is therefore rejected

FIGURE 3.29 The re-entrant processing explanation of backward masking.

SUMMARY OF SECTION 6

- There appear to be at least two partially distinct but interconnected streams of information flowing back from the retina to the primary visual cortex.
- From here, a ventral stream leads to the inferotemporal cortex and a dorsal stream to the parietal cortex.
- There is evidence that the ventral stream may be involved in perception for recognition and the dorsal stream in perception for action.
- Thus the dorsal stream would be better at dealing with the type of perception dealt with by Gibson and the ventral stream with the type of perception dealt with by Marr and the constructivist approach.
- Enns and Di Lollo's (2000) re-entrant processing explanation of backward masking was based on a combination of bottom-up and top-down perception.

7 CONCLUSION

We started this chapter by promising to show you just how complex even the perception of simple objects can be. We hope you now have some idea of these complexities and of the problems that face any potential theory of visual perception. You have also seen how rich the field of perception is. There are many influential theories that have had a profound impact on both our understanding of perception and the way we approach cognitive psychology more generally. For example, Gibson showed us the importance of considering how we interact with the real world and Marr demonstrated the advantages of the modular approach to information processing. We have also seen that although these different theories may seem contradictory at first glance, it could well be that they are all describing vital but different aspects of the perceptual process, which achieve different goals and are dealt with by different parts of the brain. So, next time you are hunting in vain for your keys, do not be too hard on yourself. Remember all the computations, descriptions, and hypotheses that your brain is having to process in order to perceive the environment around you.

FURTHER READING

The further reading suggestions below are divided into three categories: classic texts that provide more information on the key theories covered in the chapter; contemporary books dedicated to visual perception that explore the area in more depth; and books that cover perception more broadly, including modalities other than visual.

Classic texts

The three theories of perception covered in this chapter were those of Gibson, Marr, and Gregory. All three have written books considered to be classic texts in psychology that are well worth reading. Marr's 1982 book does get a little technical but also includes an important analysis of previous theories, and sets out an extremely influential approach for studying cognition. Gibson's and Gregory's books are both very readable. Sadly, Richard Gregory passed away in the year this 2nd edition was written, but his work will remain key reading for many decades to come. Indeed, *Eye and Brain* should be read by every student of cognitive psychology.

Gibson, J.J. (1986) The Ecological Approach to Visual Perception, Hillsdale, NJ, Lawrence Erlbaum Associates, Inc.

Gregory, R.L. (1997) Eye and Brain: The Psychology of Seeing, Oxford, Oxford University Press.

Marr, D. (1982) Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, New York, W.H. Freeman & Company.

Contemporary books on visual perception

Although there are more recent texts, the textbook by Bruce, Green, and Georgeson is still one of the most complete and readable accounts of visual perception.

Bruce, V., Green, P.A., and Georgeson, M.A. (2003) Visual Perception: Physiology, Psychology and Ecology, 4th edition, Hove, Psychology Press.

Books covering perception more broadly

Perception across different modalities (including hearing and touch) and also the biological properties of the brain are covered very well indeed in George Mather's excellent *Foundations of Sensation and Perception*. The edited book, *The Social Psychology of Visual Perception* by Balcetis and Lassiter, offers a good coverage of cognition and neuroscience but also integrates work on social psychology and vision to look at how factors such as goals, wishes, emotions, race, and culture affect perception.

Mather, G. (2008) Foundations of Sensation and Perception, Hove, Psychology Press.

Balcetis, E. and Lassiter, G.D. (eds.) (2010) The Social Psychology of Visual Perception, New York, Psychology Press.

REFERENCES

Aksentijevic, A., Elliott, M.A., and Barber, P.J. (2001) 'Dynamics of perceptual grouping: similarities in the organization of visual and auditory groups', *Visual Cognition*, vol.8, pp.349–58.

Atherton, M. (2002) 'The origins of the sensation/perception distinction', in Heyer, D. and Mausfeld, R. (eds.) *Perception and the Physical World: Psychological and Philosophical Issues in Perception*, Chichester, John Wiley & Sons Ltd.

ATSB (2009) Aircraft proximity event, VH-FKU and VH-HJH Bankstown Aerodrome, New South Wales. Australian Transport Safety Bureau, Aviation Occurrence Investigation, AO-2009-054.

Baizer, J.S., Ungerleider, L.G., and Desimone, R. (1991) 'Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques', *Journal of Neuroscience*, vol.11, pp.168–90.

Beck, D.M., Muggleton, N., Walsh, V., and Lavie, N. (2005) 'Right parietal cortex plays a critical role in change blindness', *Cerebral Cortex*, vol.16, no.5, pp.712–717.

Binsted, G. and Carlton, L.G. (2002) 'When is movement controlled by the dorsal stream?' *Behavioral and Brain Sciences*, vol.25, pp.97–8.

Bridgeman, B. (1992) 'Conscious vs unconscious processing: the case of vision', *Theory & Psychology*, vol.2, pp.73–88.

Bridgeman, B., Peery, S., and Anand, S. (1997) 'Interaction of cognitive and sensorimotor maps of visual space', *Perception and Psychophysics*, vol.59, pp.456–69.

Bullier, J. and Nowak, L.G. (1995) 'Parallel versus serial processing: new vistas on the distributed organization of the visual system', *Current Opinion in Neurobiology*, vol.5, pp.497–503.

Courtney, S.M., Ungerleider, L.G., Keil, K., and Haxby, J.V. (1996) 'Object and spatial visual working memory activate separate neural systems in human cortex', *Cerebral Cortex*, vol.6, pp.39–49.

Creem, S.H. and Proffitt, D.R. (1998) 'Two memories for geographical slant: separation and interdependence of action and awareness', *Psychonomic Bulletin and Review*, vol.5, pp.22–36.

Di Lollo, V., Bischof, W.F., and Dixon, P. (1993) 'Stimulus-onset asynchrony is not necessary for motion perception or metacontrast masking', *Psychological Science*, vol.4, pp.260–3.

Di Lollo, V., Enns, J.T., and Rensink, R.A. (2000) 'Competition for consciousness among visual events: the psychophysics of re-entrant pathways', *Journal of Experimental Psychology: General*, vol.129, pp.481–507.

Enns, J.T. and Di Lollo, V. (1997) 'Object substitution: a new form of masking in unattended visual locations', *Psychological Science*, vol.8, pp.135–9.

Enns, J.T. and Di Lollo, V. (2000) 'What's new in visual masking?' *Trends in Cognitive Science*, vol.4, pp.345–52.

Enns, J.T. and Rensick, R.A. (1990) 'Sensitivity to three-dimensional orientation from line drawings', *Psychological Review*, vol.98, pp.335–51.

Felleman, D.J. and Van Essen, D.C. (1991) 'Distributed hierarchical processing in primate visual cortex', *Cerebral Cortex*, vol.1, pp.1–47.

Fitts, P.M. (1964) 'Perceptual-motor skills learning', in Melton, A.W. (ed.) *Categories of Human Learning*, New York, Academic Press.

Gellatly, A.R.H., Pilling, M., Carter, W., and Guest, D. (2010) 'How does target duration affect object substitution masking?' *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1267–79.

Gibson, J.J. (1947) 'Motion picture testing and research', *AAF Aviation Psychology Research Report No. 7*, Washington, DC, Government Printing Office.

Gibson, J.J. (1950) *The Perception of the Visual World*, Boston, MA, Houghton Mifflin.

Gibson, J.J. (1966) *The Senses Considered as Perceptual Systems*, Boston, MA, Houghton Mifflin.

Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*, Hillsdale, NJ, Lawrence Erlbaum Associates.

Goodale, M.A. and Milner, A.D. (1992) 'Separate visual pathways for perception and action', *Trends in Neurosciences*, vol.15, no.1, pp.20–5.

Gregory, R.L. (1980) 'Perceptions as hypotheses', *Philosophical Transactions of the Royal Society of London, B*, vol.290, pp.181–97.

Ho, C.E. (1998) 'Letter recognition reveals pathways of second-order and third-order motion', *Proceedings of the National Academy of Sciences of the United States of America*, vol.95, no.1, pp.400–4.

Hupe, J.M., James, A.C., Payne, B.R., Lomber, S.G., Girard, P., and Bullier, J. (1998) 'Cortical feedback improves discrimination between figure and ground by V1, V2 and V3 neurons', *Nature*, vol.394, pp.784–7.

Kanizsa, G. (1976) 'Subjective contours', *Scientific American*, vol.234, no.4, pp.48–52.

Koffka, K. (1935) *Principles of Gestalt Psychology*, New York, Harcourt Brace.

Kohler, W. (1947) *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*, New York, Liveright Publishing Corporation.

Logothesis, N.K. (1994) 'Physiological studies of motion inputs', in Smith, A.T. (ed.) *Visual Detection of Motion*, London, Academic Press.

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, W.H. Freeman & Company.

Marr, D. and Hildreth, E. (1980) 'Theory of edge detection', *Proceedings of the Royal Society of London*, *B*, vol.207, pp.187–217.

Milner, A.D. and Goodale, M.A. (1995) *The Visual Brain in Action*, Oxford, Oxford University Press.

Milner, A.D. and Goodale, M.D. (1998) 'The visual brain in action', *Psyche*, vol.4, pp.1–14.

Neisser, U. (1994) 'Multiple systems: a new approach to cognitive theory', *European Journal of Cognitive Psychology*, vol.6, no.3, pp.225–41.

Norman, J. (2001) 'Ecological psychology and the two visual systems: not to worry', *Ecological Psychology*, vol.13, no.2, pp.135–45.

Norman, J. (2002) 'Two visual systems and two theories of perception: an attempt to reconcile the constructivist and ecological approaches', *Behavioral and Brain Sciences*, vol.25, no.1, pp.73–96.

Penrose, L.S. and Penrose, R. (1958) 'Impossible objects: a special type of illusion', *British Journal of Psychology*, vol.49, p.31.

Pilling, M. and Gellatly, A.R.H. (2010) 'Object substitution masking and the object updating hypothesis', *Psychonomic Bulletin and Review*, vol.17, pp.737–42.

Rao, S.C., Rainer, G., and Miller, E.K. (1997) 'Integration of what and where in the primate prefrontal cortex', *Science*, vol.276, pp.821–4.

Rock, I. (1977) 'In defense of unconscious inference', in Epstein, W. (ed.) *Stability and Constancy in Visual Perception: Mechanisms and Processes*, New York, Wiley.

Rock, I. (1983) *The Logic of Perception*, Cambridge, MA, MIT Press.

Rock, I. (1997) *Indirect Perception*, Cambridge, MA, MIT Press.

Schneider, G.E. (1967) 'Contrasting visuomotor functions of tectum and cortex in the golden hamster', *Psychologische Forschung*, vol.31, pp.52–62.

Schneider, G.E. (1969) 'Two visual systems', *Science*, vol.163, no.3870, pp.895–902.

Sedgwick, H.A. (1973) *The Visible Horizon*, Unpublished PhD thesis, Cornell University Library.

Shapley, R. (1995) 'Parallel neural pathways and visual function', in Gazzaniga, M.S. (ed.) *The Cognitive Neurosciences*, Cambridge, MA, MIT Press.

Street, R.F. (1931) *A Gestalt Completion Test*, New York, Bureau of Publications, Teachers College, Columbia University.

Ungerleider, L.G. and Mishkin, M. (1982) 'Two cortical visual systems', in Ingle, D.J., Goodale, M.A., and Mansfield, R.J.W. (eds.) *Analysis of Visual Behaviour*, Cambridge, MA, MIT Press.

Wade, N.J. and Bruce, V. (2001) 'Surveying the scene: 100 years of British vision', *The British Journal of Psychology*, vol.92, no.1, pp.79–113.

Werthiemer, M. (1923) 'Untersuchungen zur Lehre von der Gestalt, II', *Psychologische Forschung*, no.4, pp.301–50. Translated as 'Laws of organization in perceptual forms', in Ellis, W.D. (ed.) (1955) *A Source Book of Gestalt Psychology*, London, Routledge and Kegan Paul.

Young, M.J., Landey, M.S., and Maloney, L.T. (1993) 'A perturbation analysis of depth perception from combinations of texture and motion cues', *Vision Research*, vol.33, pp.2685–96.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

CHAPTER 4

RECOGNITION

Graham Pike and Nicola Brace

1 INTRODUCTION

In the last chapter on perception, we explored some of the cognitive processes involved in forming a mental description of the environment based on input from the senses. As well as being able to determine the position and shape of the objects around us, it is also possible to recognize *what* we are seeing. Unless we fully accept Gibson's concept of affordance (and it's safe to say that we don't), there must be another step: another set of processes that transform the basic descriptions of objects generated by analysing the retinal image into objects that are familiar to us and that we can recognize as specific objects, such as a mobile phone, coat, or cow.

The same is, of course, true of our other senses; for example, when we listen we may hear music, car engines, and voices. Again, there must be cognitive processes that somehow transform the auditory input of sound waves into what we recognize as an environment of voices, music, and cars.

Let's stop for a moment and consider the basic steps that might be involved in the process of visually recognizing an object:

- First, there must be processes that are able to construct an internal representation (referred to as a 'description') of the object, based on the information in the retinal image.
- Second, there must be processes that are able to store this description so that we can recognize the object if we see it again.
- Third, there must be processes that somehow compare the description of the object that we can currently see to the descriptions of objects that we have stored.
- Lastly, it is very likely that we have seen objects from many different angles, yet are able to recognize them regardless of the current angle of view. As we shall see, the nature of the mechanism that allows us to do this is an important and much debated point.

A basic diagram displaying the recognition process is provided in Figure 4.1. In one sense, the process of recognition is the process of generating and comparing descriptions of objects that are currently in view with descriptions of objects that we have seen previously. It is worth noting that this is a very simplistic way of viewing and describing recognition, and in Section 2 we shall look at some of the problems with this simplistic approach.

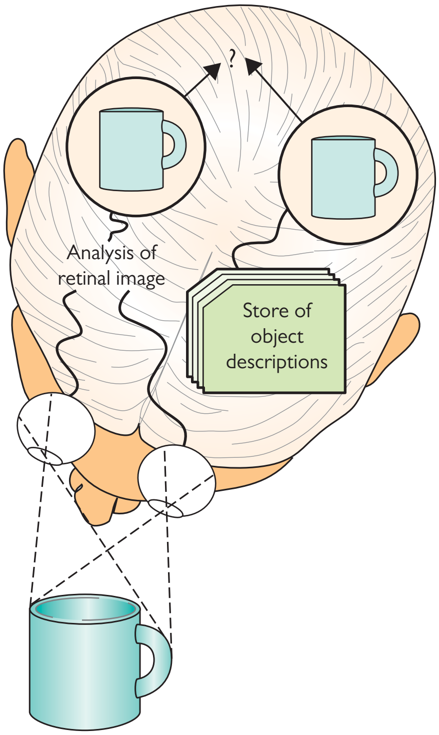


FIGURE 4.1 The basics of the recognition process.

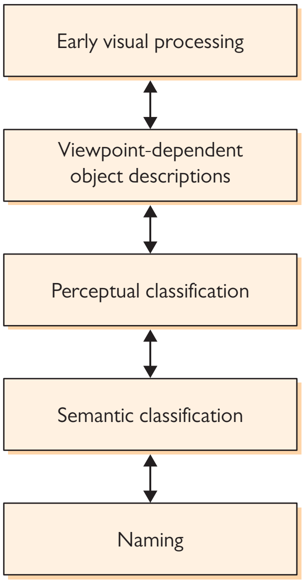

Early visual processing
Viewpoint-dependent object descriptions
Perceptual classification
Semantic classification
Naming

FIGURE 4.2 Model of object recognition suggested by Humphreys and Bruce (1989).

1.1 Recognition in the wider context of cognition

In Figure 4.2, we can see how Humphreys and Bruce (1989) summarized the way in which object recognition fits into a wider context of cognition that includes perception (perceptual classification), categorization (semantic classification), and naming. As you can see from Figure 4.2, the first stage in the process is the early visual processing of the retinal image. One example of this form of processing is that which produces Marr's full primal sketch (Marr, 1982). In the second stage a description of the object is generated, but this description is dependent on the viewpoint of the observer. This stage is therefore similar to what Marr (1982) referred to as the 2½D sketch.

Humphreys and Bruce refer to the next stage as 'perceptual classification' and it is really this stage that we have been discussing so far in this chapter. Perceptual classification involves a comparison of the information regarding the object in view with descriptions of objects that have been stored previously. It is at this stage that the object is 'recognized'.

Once the object has been recognized, or perceptually classified, it can then be 'semantically classified'. This process, also referred to as 'categorization', is examined in the next chapter. Once this stage has been achieved, the object can then be named, aspects of which will be examined in the later chapters on language.

SUMMARY OF SECTION 1

- As well as being able to determine the location and shape of an object, or the location and pitch of a sound, we also have to be able to recognize what they are.
- A basic model of recognition requires that a description from sensory input is generated and compared with descriptions stored in memory.
- Recognition must come after the initial processes of perception and before the stages in which an object can be first semantically classified and then named.

2 DIFFERENT TYPES OF RECOGNITION

As we have stated above, the view that recognition involves comparing an object description generated from the retinal image to descriptions stored in long-term memory is very simplistic. In fact there are quite different *types* of recognition, depending on what it is we are trying to recognize and how we go about trying to recognize it. Throughout this chapter we shall be exploring these different types of recognition and examining some of the issues that suggest the process of recognition is far more complex than the simplistic model presented in Figure 4.1 suggests.

2.1 Object and face recognition

The end point of Humphreys and Bruce's (1989) model of recognition (Figure 4.2) is the naming stage. Naming, of course, is not a necessary component of being able to recognize an object: even if an animal has no capacity for language, it can still recognize objects. But the names we give things do provide a clue to the fact that there are different types of recognition.

Try Activity 4.1. In completing this, you may have realized that the name you provided for the left image was the category to which the object belonged, whilst the name for the right image corresponded to an individual rather than a category (i.e. you did not name the image 'a face').

Naming reveals that it is possible to recognize objects in different ways. When we see objects such as fruit and furniture we tend to concentrate on which category they belong to, and when we provide names for them, these are usually the name for that category. Thus, we are making **between-category** distinctions such as 'that object is an orange and that one is a table'. However, when we see a face we often do more than recognize that the object belongs to the category of objects known as 'faces'; we also work out *whose* face it is. In other words, we make a **within-category** distinction.

The difference between within- and between-category recognition is one reason why **face recognition** is generally researched as a separate topic from **object recognition**. In addition, there are some issues that are unique to face recognition, such as:

- The internal features of a face can move, which changes the appearance of the face.
- This movement can serve to express emotional and social cues.
- Faces can change quite dramatically over time, due to ageing or haircuts for instance.

As well as distinguishing between face recognition and more general object recognition, it is possible to identify a number of different types of face recognition. One such distinction is between recognizing familiar and unfamiliar faces. Pike *et al.* (2000) reported that people were often able to identify E-FIT see images even when other participants had rated them as a poor likeness. However, like the E-FIT you will see in Activity 4.2, these were images of famous people whose faces would have been familiar to the participants. Considerable evidence suggests we are not so accurate at recognizing even real faces that are not so familiar to us. For example, many witnesses express uncertainty when asked to identify the perpetrator of a crime from a line-up (Pike *et al.*, 2001). Even when the anxiety of the witness is reduced by using a video identification parade, identification accuracy is far from perfect (Kemp *et al.*, 2001).

A second distinction that applies to types of face recognition is that between recognizing whose face you are looking at and recognizing what emotion it may be portraying. You can imagine that the importance of

✱ ACTIVITY 4.1

Figure 4.3 shows two images. See if you can name them.

You probably provided the names 'apple' and 'Sigmund Freud'. These are evidently two different types of name, but can you describe why these two types of name are so different? (Hint: think about how many different apples and Sigmund Freuds there are.)

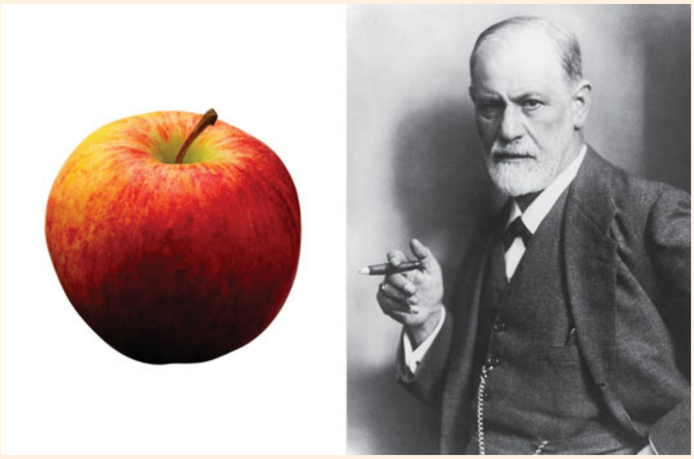


FIGURE 4.3

✱ ACTIVITY 4.2

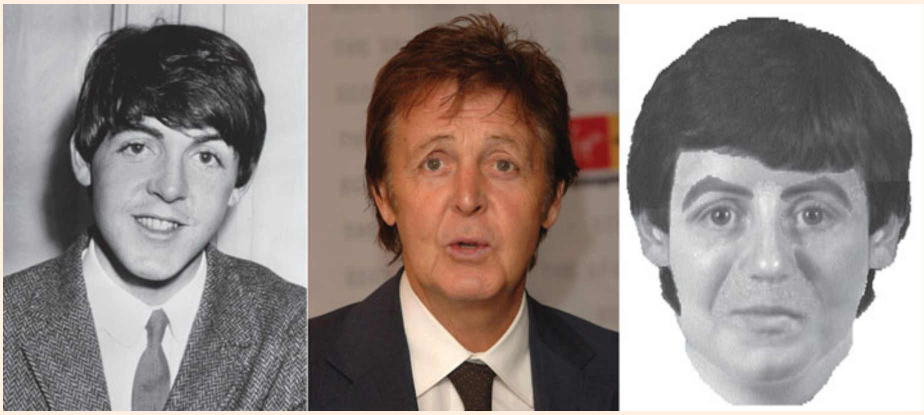


FIGURE 4.4

Can you identify the person depicted in the three images shown in Figure 4.4?

COMMENT

The images are of Paul McCartney and you were (probably) able to recognize him from all three images, even though there are some quite obvious differences in appearance. In fact, you were probably able to recognize the E-FIT image of him (right-hand image), even though this is constructed by combining together features from several other faces. So, we can recognize a face that is familiar to us even when quite large changes have been introduced.

faces in conveying emotional state and in facilitating social interactions has led us to develop some very sophisticated cognitive processes for interpreting facial expressions. In fact, we are able to judge the emotion being displayed on a face with great accuracy (the cognition involved in perceiving emotion is considered in Chapter 19) and are very sensitive to eye movements in those around us. It is tempting to think that we may have evolved a specific set of cognitive processes for recognizing faces and the emotions they express because of the social importance of this information. However, there is evidence (Young *et al.*, 1993) that although we do have specific processes for recognizing emotions, these processes are not involved in recognizing identity. We shall return to the difference between emotion and identity recognition later in this chapter, but logically you can see that you need to be able to tell whether someone is angry or happy regardless of whether you can recognize them or not. Likewise, you need to be able to recognize who someone is regardless of whether they appear happy or angry.

The question of whether faces are recognized by the same cognitive processes that are used to recognize other objects has been at the centre of a great deal of research. Although a definitive answer as to just how different face recognition is from general object recognition has yet to be provided, the two have tended to be treated as different areas of research. Because of this, we have divided this chapter into two main areas of discussion. The first (the rest of Section 2 and Section 3) will look at theories of how we recognize objects, and the second (Sections 4 to 7) will look at models of face recognition and examine in more depth the question of whether faces are recognized by special processes.

2.2 Active processing – recognizing objects by touch

One limitation of the basic recognition procedure we suggested in Section 1 is that it treats recognition as a passive process. Gibson (1986) stressed that perception is an active process and that we are beings who interact with and investigate the environment. In examining how Gibson's idea of active perception might apply to recognition, we will temporarily switch modalities from vision to touch. One reason for concentrating on touch is that purely passive object recognition through touch would be almost impossible. Although there may be some objects that you can recognize if they were simply placed on your hand, most objects would require exploration. We have evolved sophisticated processes for exploring the environment and objects using touch in very exact and careful ways.

First, we have tremendous control over our hands, so that we can both move our fingers precisely and also apply varying degrees of pressure to objects in a very measured way. This is done by employing a **feedback** system, whereby information from touch receptors allows the brain to control the location and amount of pressure applied by the fingers. As well as being able to regulate touch precisely, we can also pinpoint the location of our limbs with great accuracy via receptors inside our muscles and joints. This information about limb location is known as **kinesthesis**, and it can be combined with information from the touch receptors to guide our hands and fingers. Of particular importance are the relative positions of your fingers as they touch the object, their orientation to your hand, and the position of your hand in relation to your arm and of your arm in relation to your body. The processes that allow us to keep track of the relative locations of all our limbs are known as **proprioception**.

So, at every moment that we are touching an object, we know the exact position of our fingers (kinesthesis) and what the object feels like at that point (touch receptor information). The information gained from this combination is referred to as **haptic information** and it can be used to generate a description of an object.

Lederman and Klatzky (1987) found that there was considerable consistency in the way in which people used their hands in order to gather haptic information. They described how participants tended to use a series of exploratory procedures when investigating an object with their hands. Lederman and Klatzky (1990) went on to study these exploratory procedures in more depth and described how each particular

ACTIVITY 4.3

Ask someone to place a variety of objects within easy reach of you (you can do this yourself if you wish). Ask them to choose objects that differ in shape, texture, and weight. Close your eyes and pick up each object in turn and try to work out what it is. As you do this, try to make a mental note of the different movements that your hands make and what each movement tells you about the object.

Table 4.1 gives a list of some of the hand movements reported by Lederman and Klatzky (1987), along with the information that these exploratory procedures tend to reveal. Did you find yourself using these movements?

TABLE 4.1 The information revealed by exploratory hand movements

Movement	Information
Enclose object in hand(s)	Overall shape
Following contours with fingers	More exact shape
Lateral motion with fingers	Texture
Press with fingers	Hardness
Static contact with fingers	Temperature
Unsupported holding	Weight

Source: based on Lederman and Klatzky, 1987, Table 1, p.345

procedure could be used to derive a certain type of information that was useful for recognizing an object. For example, if shape was important in recognizing the object people tended to move their fingers around the object's contours, and if texture was important they would move their fingers across the surface of the object.

Although haptic perception can be used to recognize objects, visual recognition has the obvious advantage that it can be used for distant objects that are out of reach. It also tends to be far quicker and more accurate in processing information about shape, particularly complex 3D shape (Lederman *et al.*, 1993). But visual perception is not so useful when it comes to judging the weight and texture of an object.

So, haptic perception is a very useful source of information and can be used to recognize certain objects. The study of haptic perception also serves to demonstrate that recognition is not necessarily passive and that much can be gained from considering it as an active process. Nor is active perception limited to touch. You saw in the last chapter how your interpretation of the impossible triangle (Figure 3.24) kept changing as you visually explored the object, corner by corner.

2.3 Recognizing two-dimensional objects

Another way of distinguishing between types of recognition is according to whether the object in question is three-dimensional (3D), such as the book in front of you, or two-dimensional (2D), such as the words in front of you. The difference between 2D and 3D object recognition takes on added significance when you consider that the description generated from the retinal image will in essence be 2D, whilst most objects tend to be 3D. In fact, much of the early research conducted on recognition processes was focused on how simple, two-dimensional 'patterns' are recognized. Although it can be argued that this work tells us little about how complex, three-dimensional objects are recognized, it does serve to highlight some of the problems that are inherent in any approach to object recognition.

By far the simplest model of visual pattern recognition postulates **template matching**. This is the idea that we have a large number of templates stored in long-term memory against which we compare the patterns we come across. For example, a template would exist for every number from 0 to 9 and for every letter from

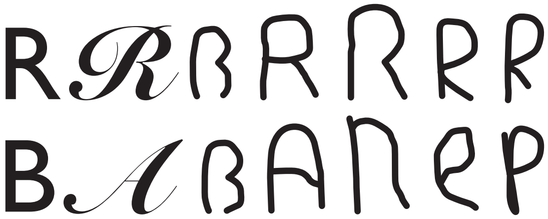


FIGURE 4.5 Different alphanumeric characters that share similar properties.

A to Z. The problem with this theory is that it cannot cope with the enormous variation in the actual patterns that are used to represent even simple things such as alphanumeric characters. For example, in Figure 4.5 the top row contains examples of the letter 'R' and the bottom row contains examples of letters that each share many similar properties with the specific example of an 'R' immediately above it. Although we do not have any great difficulty in reading these letters, it is hard to see how a simple template could be created that would accept every example in the top row as a letter 'R' and reject every example in the bottom row.

If the problem with template matching is that the template cannot deal with variation in the stimulus it has to recognize, then we have to look at some way of representing objects that is not so reliant on the exact visible pattern. One way of doing this is to try to extract the key characteristics or features of an object. In the case of alphanumeric characters, these features could be the number of curved and straight lines and the relationship between them. An 'O' might therefore be represented as a single continuous curve, a 'P' as one vertical line and one discontinuous curve, and a 'T' as one horizontal and one vertical line that form two right angles.

One of the most influential feature recognition theories is the Pandemonium system, so called because processing units known as 'demons' were used to detect each feature. This system was designed as the basis for a computer program to decode Morse code signals (Selfridge, 1959) and was later adapted by Neisser (1967) to recognize alphanumeric characters. Although Pandemonium systems have been useful in recognizing simple, highly constrained patterns, they do not provide a particularly useful model of human object recognition. A central flaw in feature recognition theories is that describing an object in terms of a list of key features does not capture the structural relations *between* features. If you look back at the feature-based descriptions provided for an 'O', 'P', and 'T' above, you will see that these three descriptions could also apply to the figures presented to the right of each letter in Figure 4.6, meaning that these shapes would be misidentified as letters.

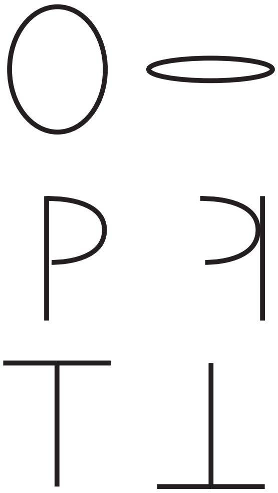


FIGURE 4.6 Examples of different patterns described by the same key features.

An approach that has had more success in explaining how both simple patterns and more complex objects might be recognized is that based on structural descriptions. Structural descriptions are made up of a series of propositions, based both on a description of the elements that comprise the object and the structural relations between them. Thus, the structural description of a letter 'L' might contain the following propositions:

- There are two lines.
- There is one horizontal line.
- There is one vertical line.
- The horizontal line supports the vertical line.
- The horizontal line extends to the right of the vertical line.
- The horizontal and vertical lines are joined at a right angle.

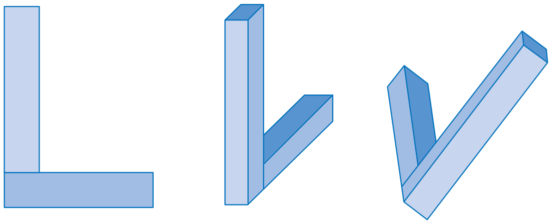


FIGURE 4.7 Three representations of a 3D ‘L’ shape.

Although the propositions stated above are expressed in language, they can be equally well expressed in other forms of symbolic representation, such as that used in a computer program.

One key advantage that structural descriptions have is that it is possible to see how they could be applied to three-dimensional objects. Consider the three representations of the character ‘L’ in Figure 4.7. Both template matching and feature recognition theories would recognize the representation to the left as being an ‘L’, but would immediately reject the other two. However, the two forms of the letter ‘L’ on the right of Figure 4.7 do share a similar structural description once we consider their three-dimensional properties.

However, in order to obtain a description that includes elements of three-dimensional structure, we must be able to turn the 2D retinal image, which is dependent on the particular view that the observer has of the object, into a 3D description that is centred not on the viewer but on the object itself. This, as you might expect, requires an even more sophisticated means of describing objects, and is the focus of the second half of Marr’s theory of vision – which we shall look at in Section 3.

2.4 Object-centred vs viewer-centred descriptions

One of the most fundamental problems in recognizing an object is that it is possible to view an object from many angles. As we have seen, any theory that treats an object as a simple pattern is likely to fail when applied to a 3D object (as with the ‘L’ in Figure 4.7). Consider writing a very simple computer program based on recognizing an object by matching patterns. As an example, Figure 4.8 contains a conceptual diagram of how a computer might be programmed to recognize a coffee mug.

But coffee cups are actually 3D objects and can be viewed from many angles. Let’s see how our simple computer program would cope if we turned our coffee cup so it was facing the other way. As you can see from Figure 4.9, the program has decided that, as the patterns do not match, the object is NOT a coffee cup.

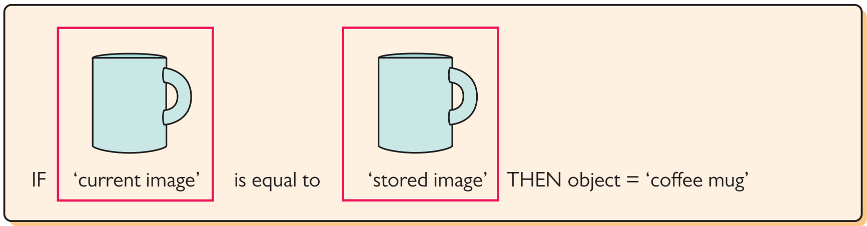


FIGURE 4.8 A simple program for recognizing an object.

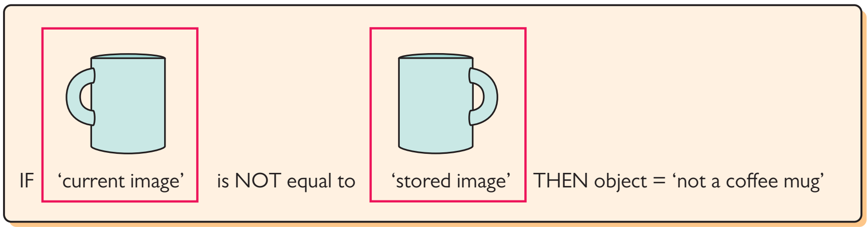


FIGURE 4.9 A simple program failing to recognize an object from a different viewpoint.

The failure of the simple program to deal with a small change in viewpoint is obviously an unacceptable flaw in any system that wishes to interact with its environment. Instead of being reliant on seeing objects from just a single viewpoint, the process of object recognition must somehow be based on descriptions of objects that allow recognition to take place *independent* of viewpoint. In fact, these processes must be tolerant of any naturally occurring change, not just changes in viewpoint. This is a very important point and one that is central to the study of object recognition.

Marr (1982) conceptualized the problem of viewpoint as that of turning the **viewer-centred description** of the object that was formed in the 2½D sketch (see Chapter 3, Section 4.3) into a **3D object-centred description** that would allow the object to be recognized despite changes in viewpoint. In the next section we shall look at how Marr suggested this might happen.

SUMMARY OF SECTION 2

- There are different types of recognition, which depend on what is being recognized and how.
- Object recognition tends to be based on making between-category distinctions and face recognition on making within-category distinctions.
- Face recognition tends to be researched apart from more general object recognition because faces can convey social and emotional information and their appearance can change.
- Recognition is not entirely a 'passive' process and can involve an active exploration of the environment. This is particularly true of haptic recognition, in which objects are recognized by touch.
- One key problem facing any theory of visual recognition is that the retinal image is essentially 2D, but objects are 3D.
- Early theories that concentrated on recognizing 2D patterns, such as template matching and feature recognition theories, are therefore not particularly useful models of human recognition.
- Theories based on abstracting a structural description of an object are better able to cope with 3D objects.
- As a 3D object can be viewed from many angles, our recognition system must be able to turn an object description centred on the viewer into one centred on the object.

3 RECOGNIZING THREE-DIMENSIONAL OBJECTS

As we saw in the previous chapter, in the first part of Marr's theory of perception, early visual processing of the retinal image eventually leads to the generation of the 2½D sketch. But the surfaces and objects in the 2½D sketch are described in relation to the viewpoint of the observer and are therefore viewer-centred descriptions. As we saw in the previous section, viewer-centred descriptions are of little use in recognizing real objects that can be seen from any angle and any distance. The second half of Marr's theory therefore concentrated on recognising objects from any angle by working out how to turn the 2½D sketch into a 3D object-centred description.

If it were not possible to generate a 3D object-centred description, the only way of accurately recognizing objects would be to store a very large number of viewer-centred descriptions. Although there are theories that have taken this approach, for now we will concentrate on the idea that recognition is best subserved by a single representation of an object that can be used to recognize it from any angle.

Marr and Nishihara (1978) suggested that objects could be represented by generating a 3D object-centred description that would allow the object to be recognized from virtually any angle. They proposed that this description was based on a **canonical coordinate frame**. This basically means that each object would be represented within a framework that was about the same shape as the object. You could imagine the representation of a carrot as being a cylinder that tapered towards one end.

This procedure appears at first glance to be somewhat paradoxical, as it would be necessary to know the approximate shape of the object before you could begin to recognize it! However, remember that the formation of the 3D object-centred description occurs after considerable analysis of the retinal image has already taken place, so some information as to the shape/outline of the object will already exist.

3.1 Marr and Nishihara's theory

Marr and Nishihara saw the first step in establishing a canonical coordinate frame as defining a central axis for the object in question. This is relatively easy to do if the object in question either has a natural line of symmetry or has a length that is noticeably greater than its width and depth (see Figure 4.10).

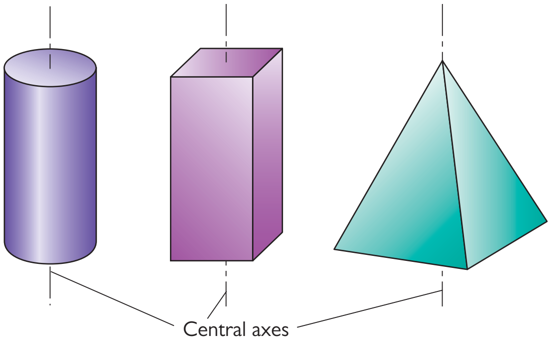


FIGURE 4.10 Locating the central axis of an object.

In fact, the generation of the central axis is so important in Marr and Nishihara's theory that it is restricted to specific objects that can be easily described by one or more **generalized cones**. A generalized cone is any 3D shape that has a cross-section of a consistent shape throughout its length. The cross-section can vary in size but not in shape. All of the objects shown in Figure 4.11 are examples of generalized cones. Although restricting the theory to generalized cones is undoubtedly one weakness of Marr and Nishihara's theory, the basic shape of many natural objects, particularly those that grow (such as animals and plants), can be described, albeit rather loosely, in this way.

To locate the central axis of an object, it is first necessary to make use of the information contained within the $2\frac{1}{2}$D sketch in order to work out what shape the object has. Marr (1977) suggested that it is possible to work out the shape of an object based on the object's **occluding contours** (these are basically the object's silhouette). The points on the object's surface

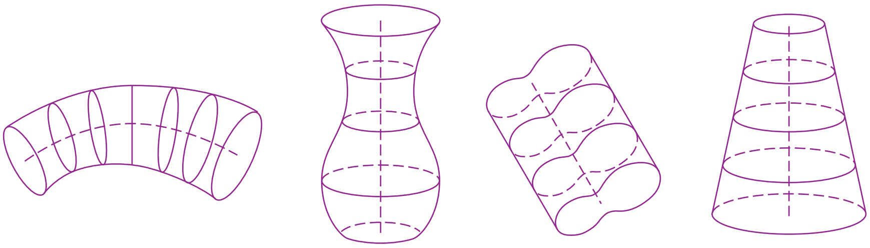


FIGURE 4.11 Four generalized cones. *Source: Marr, 1982, Figure 3.59, p.224*

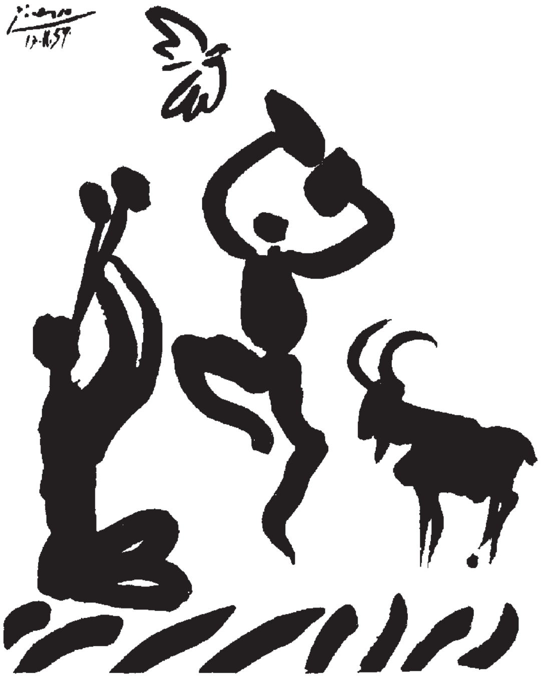


FIGURE 4.12 *Rites of Spring* by Picasso.

that correspond to the boundary of its silhouette are of particular importance in Marr's theory, and he referred to them as the **contour generator** – because they can be used to generate the contour of the object.

As Marr (1982) points out, we seem to have no problems in deriving 3D shapes from silhouettes such as those used in Picasso's *Rites of Spring* (see Figure 4.12). However, as the silhouette of an object is two-dimensional, it is possible that it could be caused by more than one 3D object. Consider the circular silhouette (a) in Figure 4.13. This could be caused by any of the 3D objects below it (if they were sufficiently rotated), yet we tend to interpret the silhouette as being produced by the sphere (b).

Marr suggested that the problem of how we can derive 3D shapes from 2D silhouettes is solved by the visual system making certain assumptions about what it is seeing. As Marr himself said, 'Somewhere buried in the perceptual machinery that can interpret silhouettes as three-dimensional shapes, there must lie some source of additional information that constrains us to see the silhouettes as we do' (Marr, 1982, p.219). Marr conceptualized this 'additional information' as coming in the form of three basic assumptions built into the computational processes:

- Each point on the contour generator corresponds to a different point on the object.

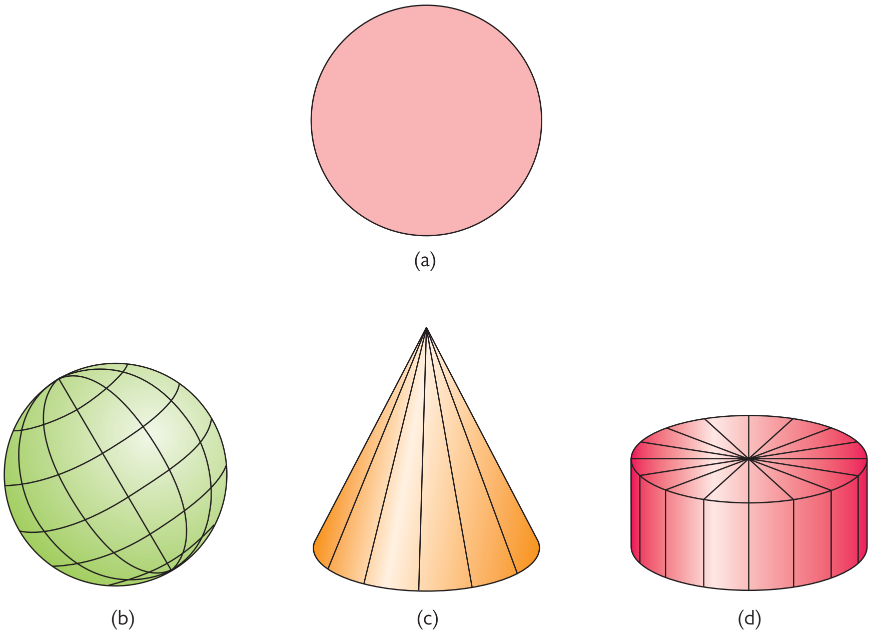


FIGURE 4.13 A silhouette (a) and three objects that could cause it (b, c, and d).

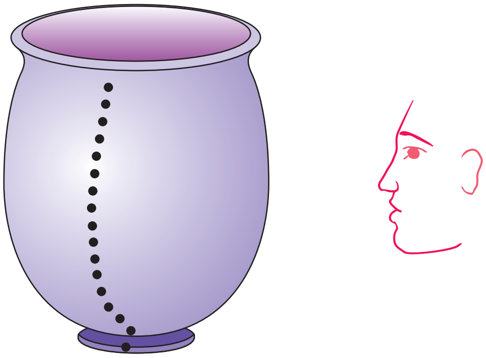


FIGURE 4.14 The black dots indicate points that lie in the same plane with respect to the viewer. *Source: Marr, 1982, Figure 3.57(d), p.220*

- Any two points that are close together on the contour in an image are also close together on the contour generator of the object.
- All the points on the contour generator lie in a single plane (i.e. are planar).

The first two points are relatively straightforward and the third assumption has been illustrated in Figure 4.14.

The third assumption, that all of the points on the contour generator are planar, is a vital component in Marr's theory, but it can be problematic. As we have seen, it is possible for two quite different objects to share the same silhouette and for the points on the silhouette to vary in their distance from the observer. We tend to interpret the contour on the left in Figure 4.15 as being a hexagon. However, this contour will be produced by the cube to the right. The problem is that the assumption that all the points on the contour are planar is violated by this view of the cube, as point (A) is further away than point (B). As the points on the cube's occluding contour are not planar, we tend to interpret its silhouette incorrectly.

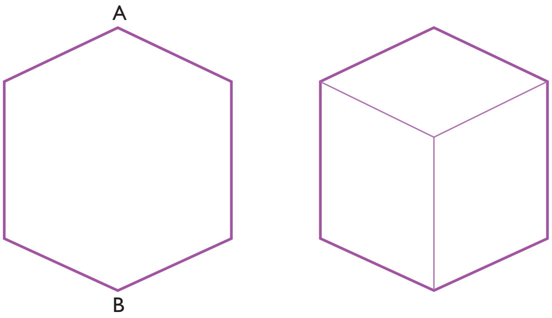


FIGURE 4.15 The contour of a cube may not be planar. *Source: based on Marr, 1982, Figure 3.58, p.221*

Once the shape of the object has been derived using its contour generator, the next step is to locate the axis/axes necessary to represent the object. It is fairly straightforward to do this when the shape is simple, as symmetry usually tells us where its axis is located, but what about more complex shapes? The answer is that we often need to represent the shape using several axes, so that the object is divided into components and one axis is used for each component (these are referred to as component axes).

In Figure 4.16, one method of locating axes suggested by Marr and Nishihara (1978) is illustrated. The object in question is a toy donkey (a). The first step (b) involves working out areas of concavity (these correspond to parts of the contour that include a bend inwards and are represented in the figure by a '–') and convexity (parts of the contour that include a bend outwards, represented by a ' + '). The shape can then be divided into sections by finding areas of sharp concavity (c) and using these to divide the object into smaller parts (d). Once the shape has been divided in this way, it is possible to represent each section via a component axis (e). These component axes can then be represented in relation to the horizontal axis of the body (f).

Figure 4.17 illustrates how it is possible to represent a quite complex object using several components, or **primitives** as Marr called them. The description of the object must allow recognition at a global level, such as being able to tell that an object is a human body, and must also incorporate more detailed information, such as the fact that a human hand has five fingers. It is therefore necessary for there to be a hierarchy of 3D models, in which each subsequent level contains a more detailed description of a specific part of the object. This means that fewer primitives will be used to represent each part of the object at the higher levels of the hierarchy.

For example, consider the description of the human body provided in Figure 4.17. At the highest level, the entire human body is described in relation to a single

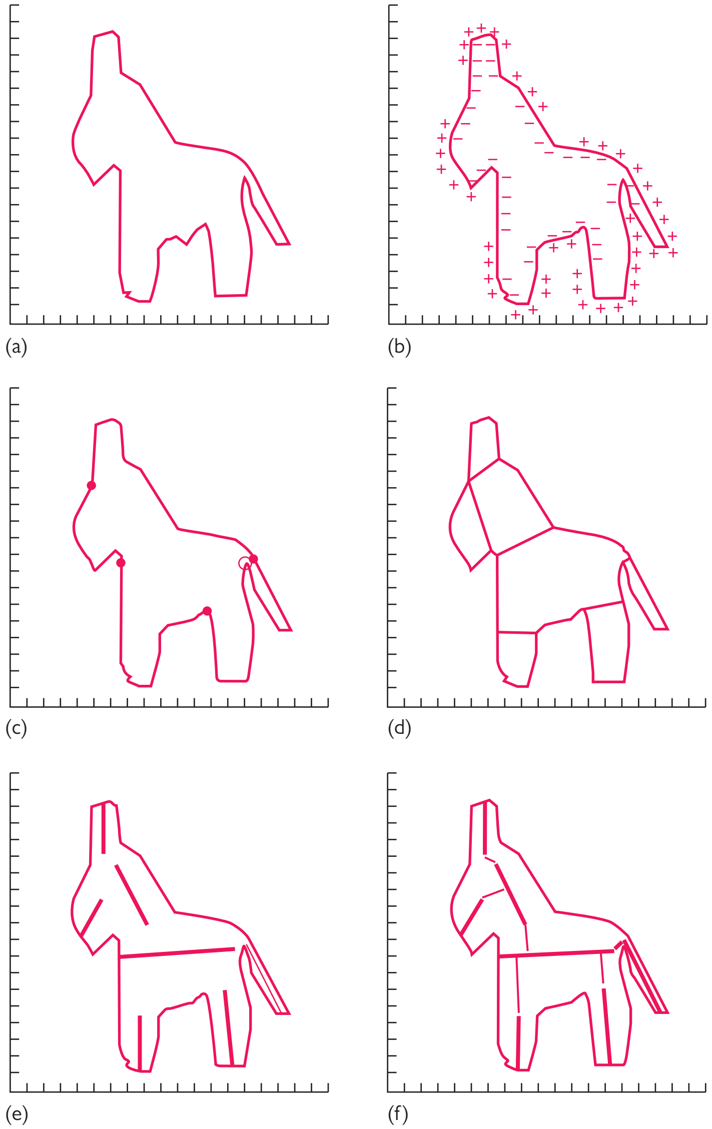


FIGURE 4.16 Locating the component axes of an object. *Source: Marr and Nishihara, 1978, Figure 6*

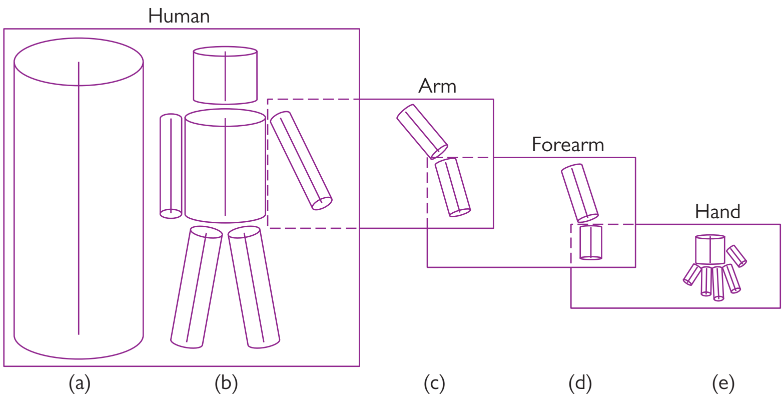

Human
Arm
Forearm
Hand
(a)
(b)
(c)
(d)
(e)

FIGURE 4.17 Marr and Nishihara's hierarchical model of a human body. *Source: Marr and Nishihara, 1978, Figure 3*

axis that runs through the centre of the body (a). This 3D model also contains the relative length and orientation of the axes that describe the head, torso, arms, and legs (b). However, no details regarding smaller parts (such as the fingers) are provided.

The axis that corresponds to each limb (b) is then used as the major axis for a more detailed description of that limb (c). For example, the axis of the cylinder representing the right arm is then used as the major axis to represent the upper and lower parts of that arm (c). The axis of the cylinder used to describe the lower part of the arm (c) is used as the major axis to describe the forearm and hand (d). Finally, the axis of the cylinder used to describe the hand (d) is used as the major axis in order to describe the five fingers (e). Thus we have a 3D model description that can be used to recognize an entire human body, as well as any of its parts.

Having derived a 3D description of the object, Marr and Nishihara (1978) saw the next step in the process of recognition as comparing this to a catalogue of 3D models, formed from the 3D descriptions of all previously seen objects. The catalogue is organized hierarchically according to the amount of detail present in the model (see Figure 4.18). At the highest level the

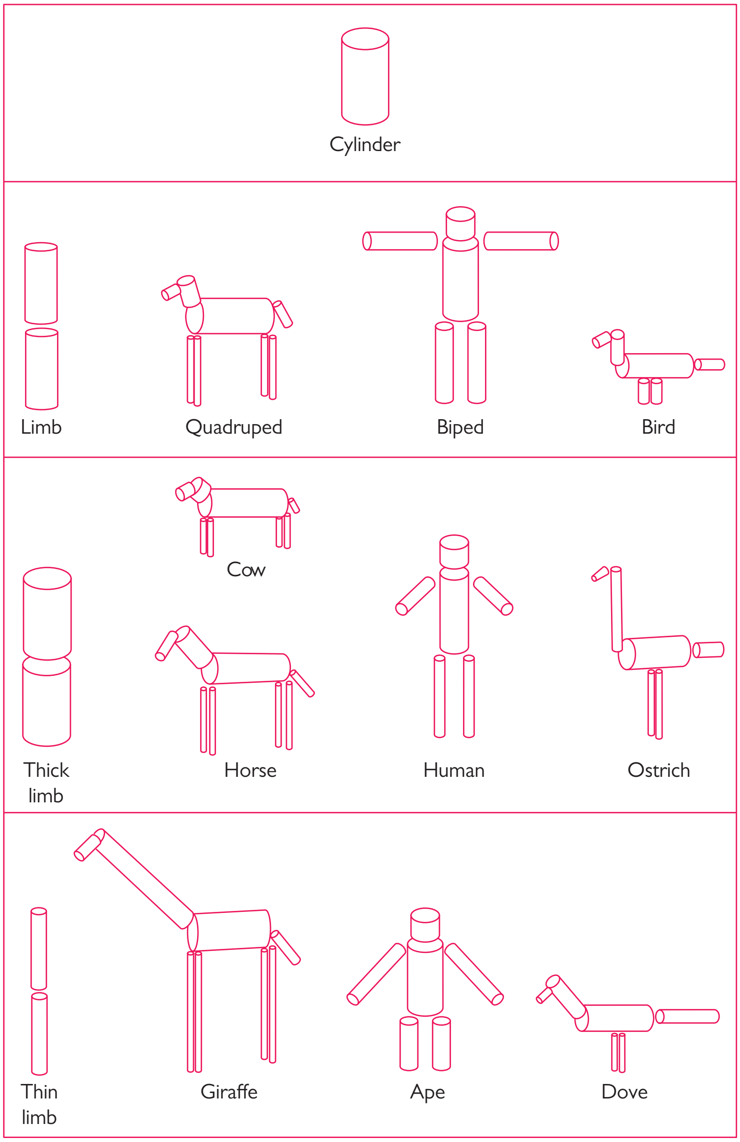


FIGURE 4.18 3D model catalogue. *Source: Marr and Nishihara, 1978, Figure 8*

catalogue consists of descriptions devoid of any decomposition into components. The next level contains more detail, corresponding to the number and basic layout of limbs as in Figure 4.17. At the next level even more detail is contained, such as that relating to the angles and lengths of component axes.

The 3D model generated of a new object (the target) is related to the catalogue, starting at the highest level. The target is compared to the stored models and the example it best matches is used as the basis for the next level of detail. The process stops when a level is reached that corresponds to the level of detail present in the target. At this point, assuming the target contains sufficient detail, a match should have been found that allows the object to be recognized.

So, the generation of a 3D model description solves several problems inherent to object recognition. As the model is 3D, it allows recognition of the object from many angles and its hierarchical nature allows recognition of the entire object whilst maintaining more detailed information about the components.

3.2 Evaluating Marr and Nishihara's theory

Although it can be difficult to study the cognition involved in object recognition, there is evidence for some of the suggestions made by Marr and Nishihara.

One of the key predictions of their theory arises from the fact that they see establishing a central axis as a vital stage in the recognition process. This means that it should be very difficult to recognize an object if it is also difficult to establish the location of its central axis. Some support for this notion comes from a study conducted by Lawson and Humphreys (1996), in which participants had to recognize objects (line drawings in this case) that had been rotated. Rotation did not appear to have an effect on recognition unless the major axis of the object was tilted towards the observer. Presumably, the disruption to recognition was due to the major axis appearing foreshortened and therefore harder to locate.

More powerful evidence in support of Marr and Nishihara's theory comes from neuropsychological case studies. Warrington and Taylor (1978) reported that patients with damage to a particular part of the right hemisphere could recognize objects when they were presented in a typical view but not when presented in an unusual view. These patients also found it very difficult to say whether two photographs (presented simultaneously) were of the same object when one image was a typical view of that object and one an unusual view.

One explanation for this effect is that the patients could not transform the two-dimensional representation of the unusual view of the object into a 3D model description. However, as well as it being difficult to establish the central axis of an object presented in an unusual view, it is also likely that rotation would cause some key features of the object to become hidden. Humphreys and Riddoch (1984) prepared images of objects in which *either* a critical feature was obscured *or* the central axis had been foreshortened through rotation. These images were presented to patients similar to those tested by Warrington and Taylor. The patients had far more problems recognizing the axis-foreshortened objects than those with a key feature hidden. The results of these studies do offer some evidence that axis location may play a central role in generating a 3D model description of an object.

3.3 Biederman's theory

Marr and Nishihara's work has been extended and adapted in several related theories of object recognition. The most influential of these was proposed by Biederman in 1987. Biederman's theory (1987a) was also based on representing complex objects using a series of more simple primitives. Unlike Marr and Nishihara, Biederman did not restrict these primitives to generalized cones. Instead he proposed that the basic building blocks for describing an object were a set of basic shapes such as cylinders and cubes, known as **geons** (an abbreviated form of the phrase 'geometric ions'). Many of these geons are generalized cones, but they also include other

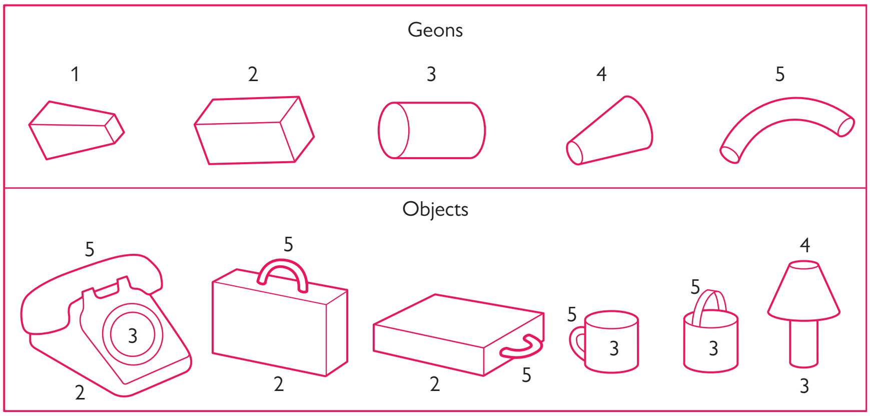


FIGURE 4.19 A selection of geons. *Source: Biederman, 1987b*

3D shapes that are very useful in representing common objects. A subset of geons is shown in the top part of Figure 4.19.

Biederman suggested that approximately 36 geons are needed in order to produce descriptions of all common objects. As with Marr and Nishihara's theory, more complex objects are represented by several different components and the division into components is based on areas of concavity.

The principal way in which Biederman's theory diverges from Marr and Nishihara's approach is the way in which a 3D description is formed from information in a 2D image – in other words, how the information in the primal sketch can be used to generate a 3D object-centred description. Biederman proposed that Marr's contour generators are not necessary to recover 3D shape, as each geon will have a key feature that remains invariant across different viewpoints. Thus, all that needs to be done is to locate these key features in the 2D primal sketch. Each feature can then be matched to a geon so that a 3D structural description of the object is generated. This description is then matched against those stored in memory.

Behind the concept of key features that remain invariant across viewpoint is the idea that some regular aspects of a 3D shape will tend to remain constant in any 2D image that is formed of that object. Biederman termed these 'non-accidental' properties to distinguish them from any regularity that was due simply to a particular viewpoint.

Biederman listed five non-accidental properties:

- *Curvilinearity* – a curve in the 2D image is produced by a curve on the object.
- *Parallelism* – lines that are parallel in the 2D image will be parallel on the object.
- *Cotermination* – two or more edges that terminate at the same point in the 2D image will terminate at the same point on the object.
- *Symmetry* – if the 2D image is symmetrical then the object will contain the same axis of symmetry.
- *Collinearity* – a straight line in the 2D image is caused by a straight line on the object.

Choosing which geon to use in order to represent an object (or part of an object) is then simply a matter of detecting these non-accidental properties and selecting a geon that shares them. For example, the 2D image of a ball will be a circle and will therefore contain no parallelism, cotermination, or collinearity, but will contain curvilinearity and an almost infinite degree of symmetry. The only geon to share these properties is a sphere, so the 3D shape of the ball is correctly described by a spherical geon.

Although these assumptions allow apparently ambiguous 2D images to be turned into an accurate 3D description, they can also lead to misinterpretation. For example, if you look at the wheel of a bicycle that is directly in front of you so that the wheel is viewed edge-on, its edges will appear to

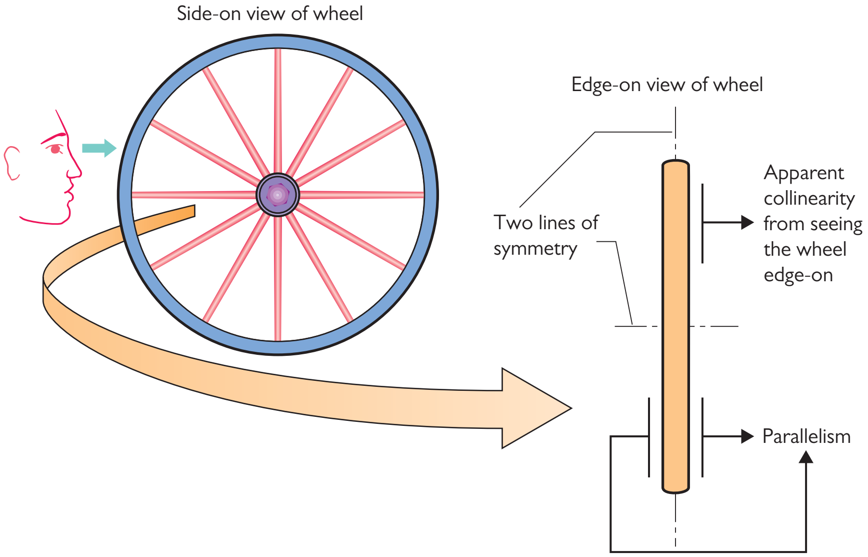


FIGURE 4.20 Apparent non-accidental properties of a wheel viewed edge-on.

have the following non-accidental properties (see Figure 4.20):

- *Collinearity* – the two vertical edges will appear as straight lines.
- *Symmetry* – there will be two lines of symmetry, one horizontal and one vertical.
- *Parallelism* – the two vertical edges will appear parallel.

However, the first of these non-accidental properties (collinearity) will be incorrect because a wheel does not contain any straight edges. We only see straight edges because of the viewpoint.

Although describing an object using non-accidental properties to select geons can lead to problems, there is evidence that supports Biederman's theory. The premise that concavities are used to divide the object into components (this premise was also used by Marr and Nishihara) was studied by presenting participants with images of objects that had part of their contours deleted. Deleting the part of the contour that corresponded to a concavity (that therefore occurred between components) resulted in a greater disruption to recognition than deleting part of the contour from elsewhere on the object (Biederman, 1987a).

3.4 Evaluating Biederman's theory and the object-centred approach

Representing objects through geometric ions or generalized cones may seem a little artificial to you, and indeed Biederman and colleagues wondered this themselves, and more specifically whether participants' sensitivity to the non-accidental properties of an object stemmed from being immersed in a manufactured world in which simple, regular shapes are very common. To test this, Biederman *et al.* (2009) compared the responses of participants drawn from two populations: students from the University of Southern California (USC) and members of the more remote tribes of the Himba, semi-nomadic people from Namibia. Both groups were asked to match a simple, regular geon to one of two other geons, which differed either according to a non-accidental property or a simple metric property. The results showed no significant difference between the responses of the Himba and the USC student participants, suggesting that the Himba are just as sensitive to the non-accidental properties of objects as are people accustomed to a more manufactured environment.

The production of an object description that is independent of viewpoint is a crucial stage in the theories of both Marr and Nishihara, and Biederman. So is there evidence that recognition does involve the generation of an object-centred description, rather then relying purely on viewer-centred descriptions?

To investigate the extent to which recognition is object-centred, Biederman and Gerhardstein (1993) used a technique known as repetition priming, where the presentation of one stimulus will make recognition of a related stimulus faster and/or more accurate. The idea behind their experiment was that if an object-centred description were being formed, then presenting one particular viewpoint of an object should facilitate (or prime) recognition of the same object presented in a different view. Their results showed that one viewpoint of an object did prime recognition of a separate viewpoint, as long as the change in the angle of viewpoint was not more than 135 degrees. However, even if the viewpoints were less than 135 degrees apart, if one or more geon(s) was hidden between the first and second view, then the amount of priming was reduced. This result supports both the idea that an object-centred description is generated (otherwise different viewpoints should not prime each other), and that this makes use of geons.

However, other researchers have reported results that do not appear consistent with Biederman and Gerhardstein's findings. Bulthoff and Edelman (1992) found that participants were generally unable to recognize complex objects that were presented in a novel viewpoint, even if the view of the object was one that should have allowed the generation of an object-centred description. Biederman and Cooper (2009) reflected on the priming work conducted since their original experiments and pointed out that although visual priming does appear to be mediated by a representation of an object's shape that is independent of its position, size, or orientation in depth, it is also the case that participants form an episodic representation of the *specific* image originally shown. That both a view-independent *and* a view-dependent representation seem to be formed by exposure to one image obviously poses a challenge to theories of object recognition that do not incorporate both forms of representation.

In the end, then, it is unlikely that recognition is completely reliant upon the generation of object-centred descriptions such as those suggested by Marr and Nishihara (1978) and by Biederman (1987a), and there may well be tasks and occasions that involve viewpoint-dependent recognition, as suggested by Tarr (1995).

One task that is hard to incorporate into either Marr and Nishihara's or Biederman's theory is that of within-category discrimination. By representing objects as models consisting of either generalized cones or geons, a wealth of information is inevitably lost. For example, it is very likely that two collie-shaped canines would be represented as identical 3D models, yet it is possible to tell a border collie from a rough collie and even to tell specific dogs apart.

It makes sense that there should be more than one way of arriving at such a complex cognitive achievement as object recognition. In the theories we have examined in this section, the process of recognition has been conceived of as almost wholly passive and based on a single retinal 'snapshot' or view. As we have stated previously, there are different types of recognition and different ways of achieving it, including taking a more active approach.

SUMMARY OF SECTION 3

- Objects can be recognized from many different angles, suggesting that the process of recognition may be based on the generation of a 3D object-centred description.
- Marr and Nishihara (1978) suggested a theory of object recognition based on generating 3D models. This was achieved by: deriving the shape of an object from the 2½D sketch; dividing it into 'primitives' using areas of sharp concavity; generating an axis for each of these components; and representing each component via a generalized cone.

- The 3D models were hierarchical in nature and so included both global and detailed information stored in a hierarchically organized catalogue.
- Biederman (1987) suggested a similar theory based on using the non-accidental properties of an object to generate a description in terms of a series of basic volumetric forms known as geons.
- Although there is evidence that supports the approach taken by Marr and Nishihara and by Biederman, there are some forms of recognition that are difficult to explain using their theories.

4 FACE RECOGNITION

The act of recognizing the face of someone familiar is part of everyday life and most of the time it happens very quickly and effortlessly. Faces are arguably our favourite visual stimuli and can provide us with a great deal of very useful information. First and foremost we are very good at determining the identity of a face that is familiar to us, but faces also provide information about gender, age, and cultural background, as well as the emotional state of a person. Although it was only in the mid-1970s that psychologists acknowledged that faces were a unique visual stimulus, since then an enormous body of research on face recognition has been conducted, meaning that in this chapter we are only able to provide you with a snapshot of some of this research.

4.1 Comparing object and face recognition

Face recognition is very problematic for the 3D model approaches we have looked at so far. If we return to Humphreys and Bruce's model of object recognition shown in Figure 4.2, we can see that theories such as those of Marr and of Biederman have concentrated on the 'perceptual classification' stage of the process. Although this stage may provide information that is useful for navigation and basic interaction with the objects we find in the environment, more complex interaction is often necessary. For example, when you are confronted by a person, you want to know not only that there is a human face in front of you, but *whose* face it is. This requires a much finer level of distinction than simply recognizing a sphere as a sphere; you must be able to tell which *specific* face is in front of you. As we shall see in Sections 4 to 7 of this chapter, the need to recognize individual faces has led to theories and research concentrating on different issues from that conducted within the area of more general object recognition.

Faces can be categorized at several different levels. At one level, we decide that the stimulus is a face as opposed to some other object. At another level, we decide that the face is female or male or derive other semantic information such as ethnic origin. We may even make attractiveness judgements. Importantly, we also decide whether the face is familiar or unfamiliar. If the face is familiar, there is also the need to decide to whom the face belongs and it is at this level that faces are rather different from other objects. It is this within-category judgement, which is like recognizing a specific cat or a specific cup, that sets face recognition apart from object recognition more generally, and it is regarded as more visually demanding because the differences between faces can be fairly minimal.

Tanaka (2001) has found evidence to liken this level of face recognition to expert recognition – for example, the expertise that certain individuals acquire through training in bird-watching or x-ray analysis. But whereas only some specifically trained people achieve object expertise, face expertise is a general expertise that we all share and acquire without specific training. Whether or not this face expertise is the result of an innate processing system or the expression of a learned skill is a matter of debate and an issue we will return to in Section 7 of this chapter.

4.2 Recognizing familiar and unfamiliar faces

So how good are we at recognizing faces and identifying people? You already saw in Activity 4.2 that it was possible to recognize a face that was familiar to you (Paul McCartney) despite quite large changes in appearance. In fact, when you think about it, you are able to recognize your family and friends from any angle, under different lighting conditions, and even when they age or change their hairstyle, and you are still likely to be able to do this in 30 years' time. There is evidence to suggest that we can remember the names and faces of school-friends over long periods of time; recognition tests revealed hardly any forgetting over a 35-year period (Bahrick *et al.*, 1975). This is not the case with all the faces we encounter though. Later work by Bahrick investigated the ability of college teachers to recognize former students taught over a 10-week period (Bahrick, 1984). The teachers had met these students three to five times a week. Although the level of correct face recognition for those taught recently was reasonably high at 69 per cent, this dropped as the number of intervening years increased so that after 8 years only 26 per cent of the former students were correctly recognized.

What about faces that are not so familiar, for example ones we've only seen once or twice before? How well do we recognize unfamiliar faces? Indeed, that might sound an odd question, as how could it be possible to recognize a face you are unfamiliar with? Of course, it is impossible to recognize a face you have never seen before, but there is a great deal of difference between recognizing a familiar face (such as those of family and friends) and recognizing a face we have only seen a few times. Most research on unfamiliar face recognition starts with a learning phase, where participants are presented with faces unknown to them that they then 'learn', and concludes with a test phase in which the participants attempt to recognize those faces they had seen previously from amongst other faces that they have never seen before. A large number of face-learning experiments have been conducted (e.g. Yin, 1969) and these have found that, when given an immediate recognition test, participants performed extremely well. (For example, Yin observed that participants correctly recognized 93 per cent of the faces previously shown to them). However, if the picture of the unfamiliar face shown in the recognition test depicted a different viewpoint or expression, then recognition rates dropped (e.g. Bruce, 1982), suggesting that what is being tested is 'recognition of a specific picture of a face' rather than 'face recognition' as we encounter it in everyday life. So, whilst we can recognize familiar faces independently of viewpoint or expression, the processing of unfamiliar faces is influenced by both viewpoint and expression. Furthermore, the environmental context in which the face is seen also has a larger influence on the recognition of unfamiliar faces than on the recognition of familiar faces (see Memon and Bruce, 1985, for a review). Finally, there is also considerable evidence from applied research on eyewitness identification decisions and from publicized cases of misidentification, such as the one involving Jean Charles De Menezes, that suggests unfamiliar face recognition is far, far more error prone than familiar face recognition.

In Box 4.1, you will see that research has also shown that even when participants do not have to recognize an unfamiliar face from memory, they are still prone to making errors.

The difference between familiar and unfamiliar face recognition might be quantitative in nature and simply reflect the amount of exposure the participants have had to the faces. However, according to Megreya and Burton (2006), it is more likely that the difference is qualitative in nature. They carried out six experiments involving face matching tasks

BOX 4.1 RESEARCH STUDY Recognizing unfamiliar faces in matching tasks

Even matching unfamiliar faces that are presented simultaneously (a task that does not test our memory) appears to be surprisingly difficult. In a field experiment, Kemp *et al.* (1997) looked at how well cashiers could match shoppers to credit cards bearing their photographs. They found that cashiers would frequently accept credit cards depicting a photograph of someone who bore a resemblance to the shopper (the correct decision rate to reject the card was only 36 per cent). Even when the photograph was of someone who bore no particular resemblance to the shopper but was of the same sex and ethnic background, the correct decision rate to reject the card was only 66 per cent (see Activity 4.4).

Other studies have demonstrated that we are not very good at matching two similar high quality photographic images when the face is unfamiliar. Bruce *et al.* (1999) showed participants a high quality video still of an unfamiliar young male target that was then presented in a line-up of similar images of nine other young men. Even when told that the target was definitely present in the line-up, participants picked it out accurately in only 80 per cent of the trials. If not told that the target was present, or if the pose of the target was varied between initial presentation and test, then performance was even worse. In fact, the performance of these participants has been matched or even exceeded by that of an automatic face recognition system tested on the same images (Burton *et al.*, 2001).

ACTIVITY 4.4

Look at the images of three faces shown in Figure 4.21. Which of the images to the left (a or b) do you think is of the same woman as that in the right-hand image (c)? These images are examples of images that were used on photo credit cards in the study conducted by Kemp *et al.* (1997).

COMMENT

The correct answer is that the left-hand image (a) is of the same woman shown in (c), but cashiers often refused to accept it due to the change in hairstyle. However, the image in the centre (b) was often incorrectly accepted as being of the woman to the right (c).

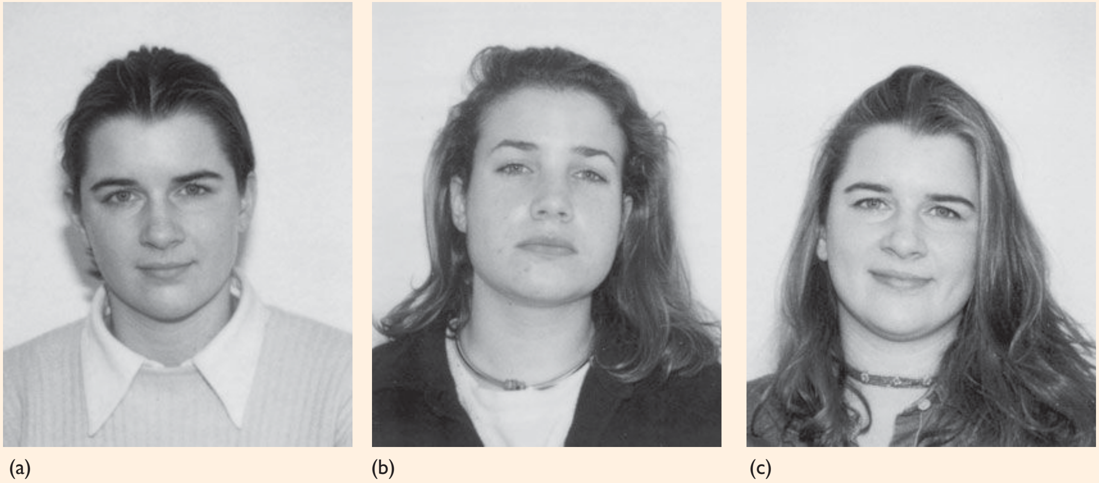


FIGURE 4.21 Three faces.

similar to those described in Box 4.1. They presented the faces in both upright and inverted (upside down) conditions, and found that whilst performance for unfamiliar faces was similar in both conditions, participants were better with familiar faces in upright conditions. This finding was replicated in a further study (Megreya and Burton, 2007). Their conclusion was that unfamiliar faces are not processed for identity like familiar faces, and whereas unfamiliar face matching may rely on image-based or pictorial processes (where the faces are treated as 'simple visual patterns'), recognition of familiar faces is likely to involve a more specialized and robust type of processing.

The next section looks more closely at what recognition of familiar faces may entail. We will start our discussion by considering some of the errors people make. These errors provide us with important information about the different systems and processes that may be involved in face recognition. Importantly, models of face recognition need to be able to account for such errors.

SUMMARY OF SECTION 4

- Face recognition is an example of a within-category judgement task.
- Our ability to identify familiar faces is extremely good and relatively unaffected by pose, viewpoint, or context.
- Recognition of unfamiliar faces is much poorer and is influenced by changes in pose, viewpoint, or context.

5 MODELLING FACE RECOGNITION

The theories of object recognition we have looked at previously centred on matching the description of an object that is in view with a stored representation. Although face recognition also involves similar matching processing, this is not usually considered the end point. In addition to matching the face we also need to access relevant semantic information and, preferably, the person's name.

5.1 Young *et al.*'s (1985) diary study

In a diary study, Young *et al.* (1985) asked 22 participants to make a record of the mistakes they made in recognizing people over an eight-week period. The recorded errors or difficulties tended to fall into different categories as shown in Table 4.2.

✱ ACTIVITY 4.5

Although we may have face recognition expertise, we do make mistakes. Before reading on, reflect for a moment and recall the last time you discovered that you failed to recognize someone you know or you mistakenly thought you recognized someone you didn't know.

TABLE 4.2 The main types of everyday errors in face recognition revealed by Young *et al.* (1985)

Types of everyday errors	Number of errors
Person misidentified	314
Person unrecognized	114
Person seemed familiar only	223
Difficulty in retrieving full details of the person	190
Decision problems	35

What do these different categories mean? '*Person misidentified*' refers to those occasions when someone unfamiliar is misidentified as someone familiar and '*Person unrecognized*' refers to occasions when someone familiar was thought to be someone unfamiliar. Both may arise because of poor viewing conditions (i.e. it is a bit dark) or because we know the person only slightly. '*Person seemed familiar only*' refers to those occasions when you recognize someone as being familiar but no other information comes to mind immediately, and '*Difficulty in retrieving full details of the person*' refers to occasions when only some semantic information is retrieved and not, for example, their name. These errors often occur when the familiar person is seen outside the context in which they are usually encountered. Finally, '*Decision problems*' refers to those occasions where you think you recognize the person but decide it cannot be them, perhaps because you believe they are currently in another country.

These errors demonstrate that it is possible to retrieve previously learned semantic information about a person without recalling their name, but that it is very unlikely that we will recall their name without also retrieving relevant semantic information. However, before we can recall either semantic information or a name, we must realize the face is familiar.

5.2 A cognitive model of face recognition

These findings on everyday errors are consistent with the notion that the recognition of familiar faces involves a sequence of processes using different types of information. Hay and Young (1982), Young *et al.* (1985), and then Bruce and Young (1986) refined a cognitive theoretical framework or model of person recognition involving such a sequence of stages. On meeting people we encode their faces. This encoded information may activate **face recognition units** (FRUs) that contain stored information about the faces we are familiar with. If there is a reasonable match between what has been encoded and what is stored in the recognition unit, then the recognition unit will be activated and allow access to semantic information about the person's identity, such as their occupation, stored in **person identity nodes** (PINs). It is only once the PIN for a face has been activated that their name can be generated. A **cognitive system** is also involved, as the information provided by the recognition system must be evaluated. As the diary study above indicated, errors in face recognition can arise because of decision problems. For example, if we know that the person doesn't live or work nearby, that knowledge may override what our recognition system is telling us and hence we may doubt that we have correctly identified the person.

So, when might the encoded information activate FRUs? As mentioned in Section 4, familiar face recognition is extremely good and relatively unaffected by pose, viewpoint, or context, and to achieve this Bruce and Young suggested that structural codes are formed over time. These codes contain the information that allows a face to be recognized despite changes in pose, expression, etc. Importantly, structural codes allow us to decide whether or not a face is familiar. The first stage of the model involves the creation of these codes resulting in two types of representations that serve different purposes. Viewer-centred descriptions result in viewpoint-specific representations and, as can be seen in Figure 4.22, these are used for expression analysis and to help make sense of speech. Expression-independent descriptions result in separate and more abstract expression-independent representations, which are used for recognition. Therefore, a face needs to be familiar if it is to activate a FRU; if it is unfamiliar then its representation will be linked to the particular view of the face that is being shown in that particular encounter.

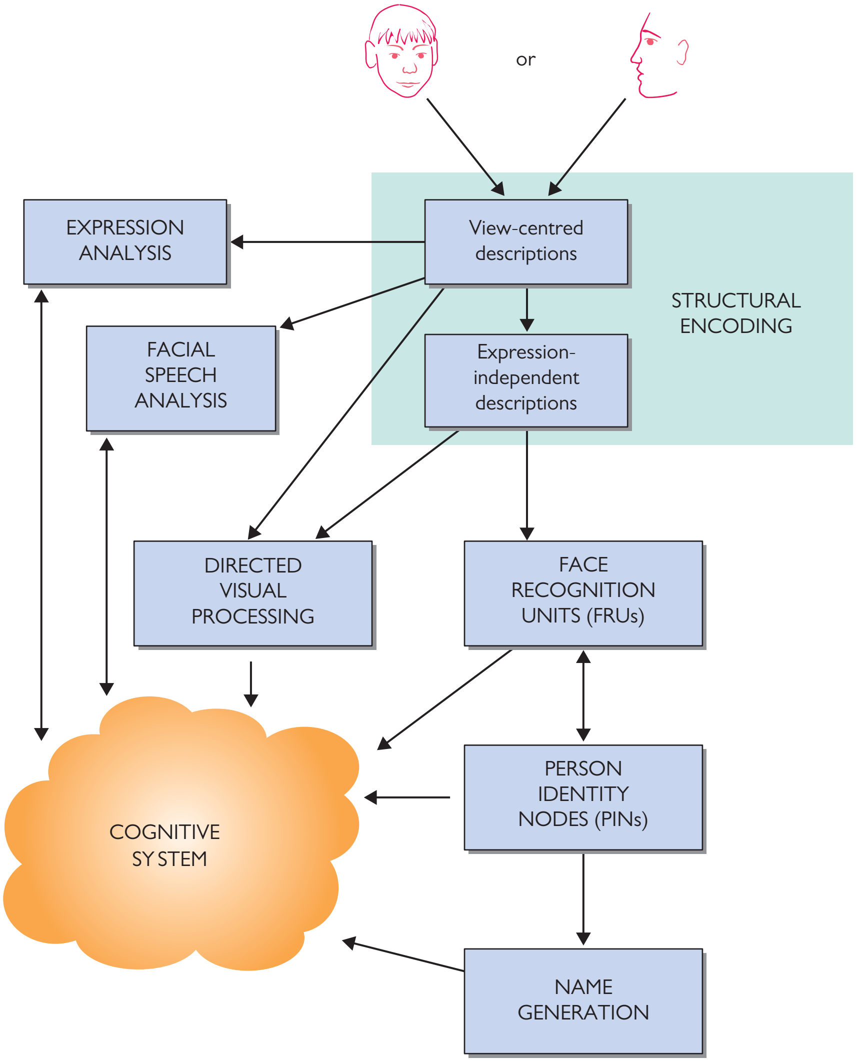

or
EXPRESSION ANALYSIS
View-centred descriptions
STRUCTURAL ENCODING
FACIAL SPEECH ANALYSIS
Expression-independent descriptions
DIRECTED VISUAL PROCESSING
FACE RECOGNITION UNITS (FRUs)
COGNITIVE SYSTEM
PERSON IDENTITY NODES (PINs)
NAME GENERATION

FIGURE 4.22 Bruce and Young's functional model for face recognition. *Source: Bruce and Young, 1986, p.312*

The Bruce and Young (1986) functional model for face recognition is presented in Figure 4.22. As you can see, there are separate routes for facial expression analysis, facial speech analysis, and face recognition; and face recognition progresses through a sequence of stages from FRUs to PINs to name generation.

The notion that different types of information are sequentially accessed is also supported by the results of experiments conducted in the laboratory. For example, Hay *et al.* (1991) showed participants 190 images of the faces of celebrities (including people likely to be familiar to the participants and those likely to be unfamiliar) and asked them to decide whether or not each face was familiar and to state the person's occupation and name. Participants did not retrieve a name without also being able to name the occupation, thus supporting the notion that semantic 'person identity' information is retrieved *before* the person's name. Other studies (e.g. Johnston and Bruce, 1990), looking at how quickly we can complete a particular task, have shown that faces can be classified as familiar more quickly than they can be classified by occupation, and furthermore that classifications that require accessing the person's name take longer than classifications involving a person's occupation or other semantic properties. These findings support the notion that perceptual classification, judging the familiarity of a person, takes place *before* semantic classification and that a person's name is accessed last. They also provide a nice demonstration of how the findings from the laboratory may support those derived in a more everyday study of face recognition, such as Young *et al.*'s (1985) diary study.

5.3 A connectionist model of face recognition

The **IAC model** (e.g. Burton *et al.*, 1990; Burton and Bruce, 1993) is a connectionist model (recall the discussion of connectionism in Chapter 1) of face recognition and an extension and implementation of the Bruce and Young model described above. IAC stands for 'interactive activation and competition network'. As this model is a computer simulation of face recognition it has been tested by seeing how compatible it is with the available evidence, and by looking at the predictions it generates.

The model comprises units that are organized into pools (see Figure 4.23). These pools contain:

- *FRUs (face recognition units):* For every familiar person, there is one FRU in the model. These are

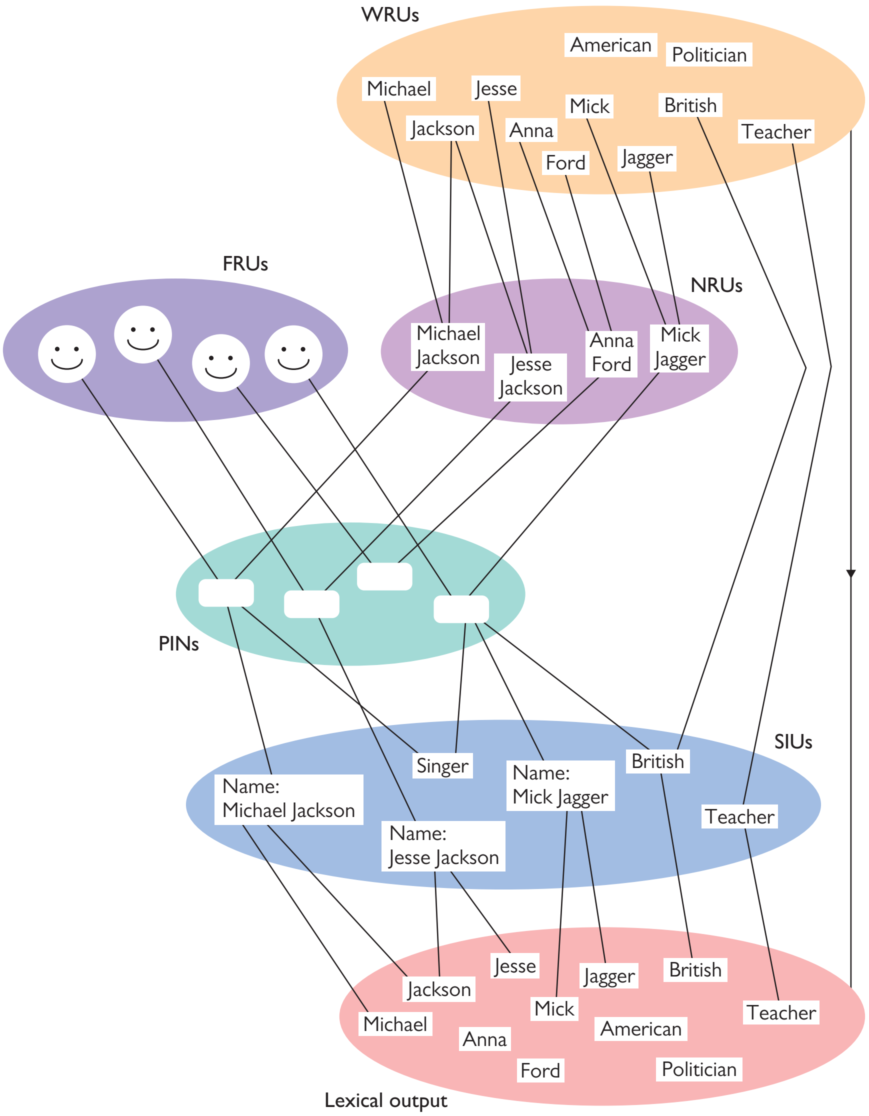

WRUs
American
Politician
Michael
Jesse
Mick
British
Jackson
Anna
Teacher
Ford
Jagger
FRUs
NRUs
Michael Jackson
Jesse Jackson
Anna Ford
Mick Jagger
PINs
SIUs
Singer
Name: Michael Jackson
Name: Mick Jagger
British
Name: Jesse Jackson
Teacher
Jesse
Jagger
British
Jackson
Mick
Teacher
Michael
Anna
American
Ford
Politician
Lexical output

FIGURE 4.23 The central architecture of the IAC model.

view-independent and seeing any recognizable view of a face will activate the appropriate FRU. These representations allow perceptual information to be mapped onto stored memories. (This is basically what was suggested in the Bruce and Young model.)

- *PINs (person identity nodes):* This is where a face is classified as belonging to a person, and there is one unit per known person.
- *SIUs (semantic information units):* Relevant semantic information is stored here, e.g. occupational category.
- *Lexical output:* Units representing output as either words or a name.

The IAC model also includes a route based on word recognition. The pool of WRUs (word recognition units) represents an input lexicon containing both specific names and more general information, such as nationality or occupation. Words that are names have direct links to a pool of NRUs (name recognition units), which are linked to PINs in the same way as FRUs. The WRUs that do not correspond to names are linked to SIUs.

Figure 4.23 shows how the pools are connected. The input systems (FRUs and NRUs) join to a common set of person identity nodes (PINs) and these are linked to units containing semantic information (SIUs). Each of the pools is illustrated here with just a few examples of the units they might contain. Many SIUs will be shared and here many people will be represented with such information as 'teacher' or 'British'.

Recognizing a face is modelled in the following way: seeing a face will activate an FRU, which in turn increases activation in the relevant PIN. As PINs are linked to SIUs, activation of the PIN will bring about activation in the relevant SIU. The notion that different types of information are sequentially accessed is therefore still present in this connectionist model. If a certain threshold is achieved in the PIN, then this signals familiarity. An important point to note is that different types of information come together at the PIN stage, including information from recognition systems specialized for faces, as well as those specialized for the recognition of written or spoken names, and familiarity is judged on this pooled information.

We mentioned before that IAC stood for 'interactive activation and competition network'. The 'interactive activation' arises from the links between units in different pools, which are *excitatory*: the FRU for Mick Jagger's face excites or activates the PIN for Mick Jagger, which in turn excites semantic information units for the name 'Mick Jagger', the occupation 'singer', and the nationality 'British'. These excitatory links are bidirectional so that excitation also runs in the opposite direction from 'singer' to Mick Jagger's PIN and Mick Jagger's FRU. However, within each pool links between units are *inhibitory* (these links are not shown in Figure 4.23), so this is where 'competition' arises. Excitement in the FRU for Mick Jagger will inhibit activity in the other FRUs, just as excitement in Mick Jagger's PIN will inhibit activity in the other PINs and excitement in one SIU will inhibit activity in another SIU. But the SIU for Mick Jagger (which might be 'singer') will also excite many other PINs (in this example, those belonging to all other singers). This means that activation of PINs will not be limited solely to the specific person in question, but some activation will also occur for anyone who is semantically related (e.g. shares the same occupation). Thus, the model incorporates the results of experiments that have shown priming effects – that you are quicker to recognize Bill Wyman if you have already seen Mick Jagger. Generally, the strength of this connectionist model is that it can account for findings from laboratory studies as well as for the everyday errors described by Young *et al.* (1985).

To conclude this section, it is important to highlight that the Bruce and Young model has been very influential and still remains so today, providing a framework for organizing the different elements involved in face recognition. It has been superseded by the IAC model, which provides more information on how FRUs and PINs actually operate and accounts for the effects of priming. Next we shall consider how well the IAC model can account for findings from neuropsychological studies.

SUMMARY OF SECTION 5

- Everyday errors suggest that recognizing faces involves sequential access to different types of information.
- A cognitive model of person recognition involving such an idealized sequence of stages has been developed (Bruce and Young, 1986).
- IAC is a connectionist model of face recognition that is an extension and implementation of this cognitive model.

6 NEUROPSYCHOLOGICAL EVIDENCE

Prosopagnosia, the inability to recognize faces whilst maintaining the ability to recognize other objects, is a well-documented phenomenon. However, cases of 'pure' prosopagnosia are exceptionally rare. It is more common to see deficits affecting other visual categories too. The recognition of all familiar faces is affected, regardless of their semantic categories (so it is not the case that the failure to recognize a face is restricted to faces of celebrities or politicians). However, as recognition from other cues, for example voice, usually remains unaffected, the condition is specific to visual recognition of faces and is not a more general impairment of the recognition of personal identity. Also, the ability to distinguish between faces is often preserved.

In this section, we shall focus on two key findings that have emerged from investigations of prosopagnosia: first, that identification of expression appears to be independent from face identification; and second, that face recognition and awareness of face recognition might also be independent of one another. It is possible that although prosopagnosics are unable to recognize faces consciously or overtly, certain types of non-conscious response may be preserved. We shall examine how the IAC model may account for this.

6.1 Separate routes for identification and emotional expression

As mentioned in Section 5, models of face recognition have proposed a route for face identification that is independent of emotional expression, and this independence has received support from experimental work and from neuropsychological research. In many cases of prosopagnosia, the ability to recognize facial expressions may be unaffected. Young *et al.* (1993) looked at ex-servicemen with unilateral brain injuries and tested familiar face recognition, unfamiliar face matching, and analysis of emotional facial expressions. Analysis of accuracy data showed evidence of selective impairments in each of these three abilities. For example, one participant with a right hemisphere lesion was selectively impaired in identifying familiar faces, whereas a different participant, also with a right hemisphere lesion, had problems only with matching unfamiliar faces. A number of other participants with left hemisphere lesions were only impaired on the facial expression tasks. Response latency data also supported the notion of a selective deficit of facial expression processing, but suggested that impairments of familiar face recognition and unfamiliar face matching were not entirely independent from one another. The findings from this study thus provide strong support for the notion that facial expression analysis and face identification seem to proceed independently of each other (and also some support for the notion that the ability to recognize familiar faces and to match unfamiliar faces may be selectively and independently impaired).

6.2 Covert face recognition

Previously, when describing models of face recognition, we did not draw a distinction between face recognition and awareness of recognition. However, neuropsychological research on prosopagnosia suggests that the distinction is important. Bauer (1984) monitored changes in autonomic nervous system activity via changes in **skin conductance response (SCR)**. These changes signal an affective or emotional reaction (you may remember reading in Chapter 2 on attention how a closely related response, GSR, was measured to look at unconscious processes). Bauer showed LF, a participant with prosopagnosia, a face and read out a list of five names, whilst simultaneously measuring SCR. If LF was asked to pick the correct name he performed no better than expected by chance. In other words, LF was overtly unable to recognize familiar people from their faces. However, LF showed a greater SCR when the correct name was read aloud compared with the incorrect names. Thus, LF was showing an affective or emotional response, but this response was not a conscious one. The term **covert recognition** is used to describe this non-conscious recognition or emotional response to the faces.

Since Bauer's work, many studies have investigated covert recognition and the issue is not whether this type of face recognition exists but how to interpret it. Bauer proposed that separate neural pathways are responsible for two independent routes to recognition, one for conscious overt recognition and one for non-conscious covert recognition. Although questions remain over exactly how overt and covert recognition processes are mediated and how these processes normally become integrated, there is support for the involvement of the two major neural pathways (see Box 4.2).

A different issue is whether those individuals who retain covert recognition can be helped to overcome their disorder. Could covert recognition be turned into overt recognition? Sergent and Poncet (1990) were the first to demonstrate such provoked overt recognition. In their study, PV was shown eight faces of famous people from the same semantic category and she was unable to identify them. However, when she was told that they all had the same occupation and she looked at the faces again, she was able to say that they were all politicians, name seven of the people, and recall biographical information about the eighth person. This and other later studies (e.g. Diamond *et al.*, 1994) have shown that provoked overt recognition can occur under certain experimental conditions, and this provides some hope for rehabilitative work.

Can the IAC model accommodate the pattern of deficits described here? Covert without overt recognition is explained in terms of attenuation (or weakening) in the connections between the FRUs and PINs. This means that when a face is seen and the FRU is activated, the weakened FRU-PIN connection strength means that excitation of the corresponding PIN is not raised above the threshold for the face to be recognized overtly. However, this weakened activation may be sufficient to raise the excitation of the PIN above its resting level, mediating covert recognition. Provoked overt recognition is explained in the following way. Telling PV that the faces are related is equivalent to strengthening the PIN-SIU connections. Unlike FRU-PIN connections, PIN-SIU links are assumed to remain intact in instances of prosopagnosia where covert recognition is observed. Once these connections are strengthened, activation is passed back from the shared SIUs to the relevant PINs. These then achieve threshold and the faces are recognized overtly. Simulations with the model confirmed this particular prediction – provoked overt recognition was successfully modelled (Morrison *et al.*, 2001).

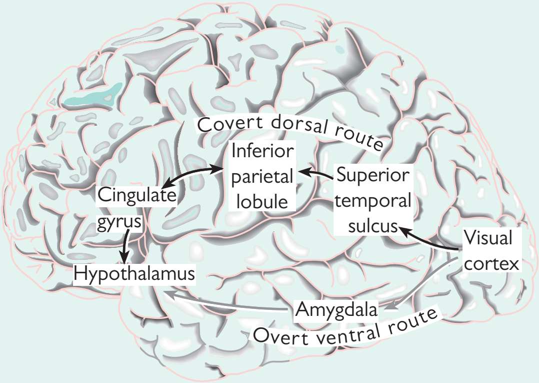
BOX 4.2 RESEARCH STUDY Capgras delusion

Capgras delusion usually occurs as part of a psychiatric illness, although it can result from brain injury. A person with Capgras delusion believes firmly that someone they know, usually a relative or close friend, has been replaced by an impostor, double, robot, or alien. Sometimes the delusion relates to objects; for example, the sufferer may believe that tools, ornaments, or other household objects have been replaced by doubles. Face and object Capgras delusion do not usually co-exist, and the disorder tends to be specific to one domain. The key point here is that individuals with a face Capgras delusion recognize a face but simultaneously refute its authenticity. Exactly why those with Capgras delusion adhere to the belief that the person must be an impostor is still being debated.

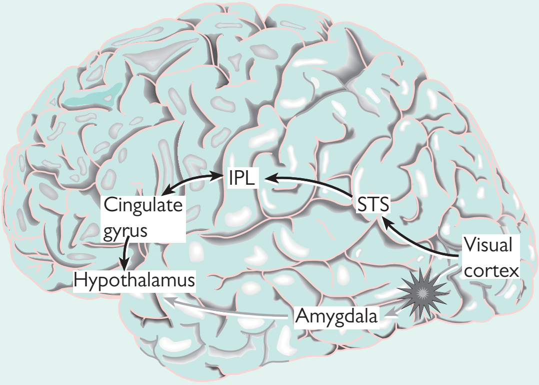
One of several explanations was offered by Ellis and Young (1990). They suggested that Capgras delusion may be a 'mirror image' of the impairments underlying prosopagnosia. Bauer (1984) proposed that the neuroanatomical pathway involved in overt recognition was the 'ventral visual-limbic pathway', whereas the pathway involved in covert recognition was the 'dorsal visual–limbic pathway'. Ellis and Young suggested that the Capgras delusion resulted from damage to such a dorsal route, so that sufferers would recognize the familiar person but not receive supporting affective information. Their prediction that individuals with Capgras delusion would recognize familiar faces but would fail to show an autonomic emotional response to these familiar faces has received support from several studies (e.g. Hirstein and Ramachandran, 1997). Whilst overt recognition is intact, covert recognition seems to be impaired.

Figure 4.24 shows normal face processing (a), with a darker arrow showing the covert dorsal route and a lighter arrow the overt ventral route. In prosopagnosia (b) the overt ventral route is thought to be damaged, and in Capgras delusion (c) the covert dorsal route is thought to be damaged.

(a) Normal face processing


Covert dorsal route
Inferior parietal lobule
Superior temporal sulcus
Cingulate gyrus
Visual cortex
Hypothalamus
Amygdala
Overt ventral route

(b) Prosopagnosia


IPL
Cingulate gyrus
STS
Visual cortex
Hypothalamus
Amygdala

(c) Capgras delusion

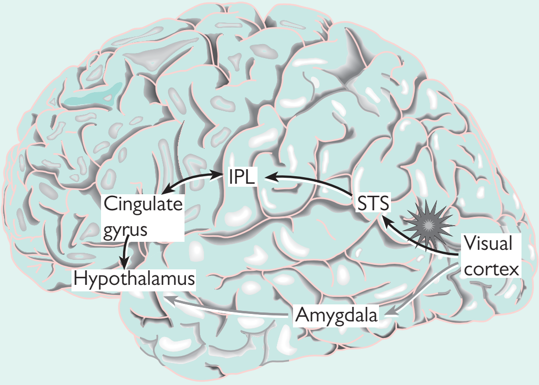

IPL
Cingulate gyrus
STS
Visual cortex
Hypothalamus
Amygdala

FIGURE 4.24 The dorsal and ventral routes in normal face processing (a), prosopagnosia (b), and Capgras delusion (c). *Source: Ellis and Lewis, 2001, Figure 3, p.154*

6.3 Evaluating the IAC model

As a model of face recognition, the IAC model is impressive in that it can account for a wide range of data from studies on face recognition. Whilst there are other models of face recognition, some of these are based on a narrower range of evidence; for example, they may have sought only to account for the findings from neuropsychological studies. As we have seen here, IAC is compatible with everyday, laboratory, and neuropsychological findings.

Hole and Bourne (2010) suggest that the weakness of IAC is that it says little about what information is extracted from the face in the first place to allow it to become familiar. How is a face encoded to allow it to be matched to an internal representation so that FRU activation then follows? One issue that has been raised is whether structural codes are indeed used to combine the descriptions formed by multiple encounters with a (familiar) face, so that what is stored in memory is an abstracted representation of the face. Longmore *et al.* (2008) have challenged such a notion after carrying out six experiments where participants learned to recognize new faces to near 100% accuracy (they termed this familiarized face recognition). Their findings revealed that when participants learned multiple views of the same face, their performance decreased when shown novel views of that face, suggesting that the different views of the same face were stored as a collection of separate images rather than as a combined representation. Therefore, what was stored appeared to be based on view-specific rather than structural codes. We shall consider other findings regarding the information that is extracted from a face in the final section of this chapter.

SUMMARY OF SECTION 6

- Prosopagnosia is the inability to recognize faces, although expressions and other objects may still be correctly identified.
- Covert face recognition, shown by autonomic responses to faces, may however be spared.
- Overt conscious face recognition and covert non-conscious face recognition are different types of face recognition that may be mediated by different neural pathways.
- Capgras delusion may be a mirror image of prosopagnosia in terms of which system remains intact and which system is damaged.
- Provoked overt recognition has been achieved in some studies and has been successfully modelled using the IAC model.

7 ARE FACES 'SPECIAL'?

In this last section we return to the issue of the difference between face recognition and object recognition, and in particular to face expertise and how we are able to discriminate so readily between faces. There are several important issues that the literature has addressed:

1. Is there a neuroanatomical location that underlies face processing and, if so, does this mean that face processing is unique and qualitatively different from the processing of other types of visual stimuli?
2. Is face processing an innate or learned skill? Have we developed a face expertise because of constant exposure to faces and practice at differentiating between them, or is this an innate ability?

3. How important are the individual features of the face, the relationships between the features, or the three-dimensional structure? Do we process the individual facial features or the face as a whole?

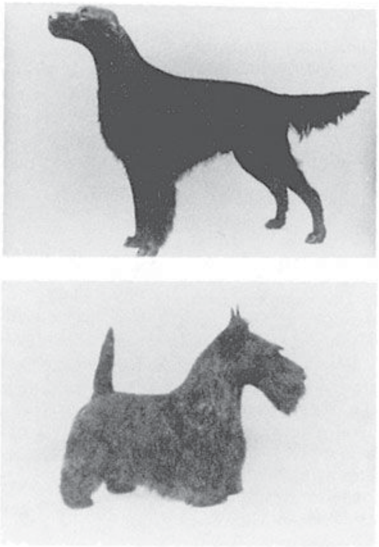
In the previous section, we looked at the syndrome of prosopagnosia and found that research implicated several neurological pathways. Of particular interest is that prosopagnosia can leave object recognition relatively intact and, in turn, face recognition has been spared in cases where object recognition has been impaired (a double dissociation). Studies using the technique of functional magnetic resonance imaging (fMRI) have found facial stimuli to activate an area in the fusiform gyrus in the posterior temporal lobes (especially in the right hemisphere), whilst non-face objects activated a different area. There is also the observation of cells specialized for faces within the monkey temporal lobe – these cells respond selectively to faces of humans and/or monkeys but not to other stimuli (e.g. geometrical shapes and bananas). There is, therefore, evidence to suggest that the processing of faces is mediated by specific areas of the brain; that there is cortical specialization for faces. But does this mean that face recognition is *unique*, that the processes used for recognizing faces are qualitatively different from those used for recognizing other visual stimuli?

There is support for the notion that there is a special mechanism from birth for processing facial information, as newborn babies show a preference for face-like visual patterns. Rather than an innate neural mechanism that processes faces, Johnson and Morton (1991) suggested that there is a mechanism that makes newborns attentive to faces, and this innate attentional bias then ensures that any system for learning visual stimuli receives a lot of face input and learns about the individual characteristics of faces. Although there is a 'kick-start' mechanism that gives face processing in newborns a special status, which serves to guide subsequent learning, soon other processing systems will come into play that may or may not be unique to faces.

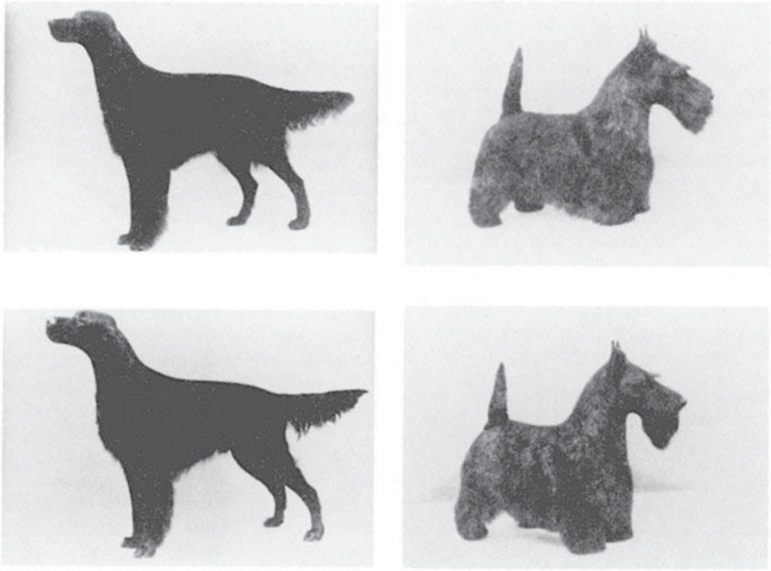
One further reason to think that face recognition is a special type of recognition, distinct from other object recognition, is that faces all tend to look alike in that they have similar features in similar positions. Given this similarity, it could be that we have to make use of a different form of visual information to recognize a face from that used to recognize, for example, a table. Some evidence that this is indeed the case comes from studies that have demonstrated that inverting, or turning upside down, visual stimuli disproportionately impairs our ability to recognize faces compared with our ability to recognize objects. This is known as the **inversion effect**. Yin (1969) and other studies since (e.g. Johnston *et al.*, 1992) have shown that inverting a photograph of a face disrupts recognition more than does inverting an image of an object. Yin looked at the influence of inversion on faces and other stimulus material including houses and aeroplanes. Although recognition memory was better for upright faces than for other material, when the stimuli were turned upside down, recognition for faces was worse than that for other material. The key question is whether this peculiar reversal of recognition accuracy for faces (from best upright to worst inverted) supports the notion that faces are processed differently from other stimuli or whether there is an alternative explanation.

Diamond and Carey (1986) investigated an alternative hypothesis, namely that the effect of inversion on faces was a result of our perceptual mechanisms becoming 'tuned' to seeing this special type of visual stimulus in the usual upright orientation. This 'tuning' or expertise would then be 'lost' when we see them inverted. Their research considered whether the inversion effect was indeed specific to human faces or whether it would in fact arise when using any class of visual stimulus with which we have a large amount of experience. To investigate this, Diamond and Carey selected participants to include both people who were not interested in dogs and people who were dog experts (mainly dog-show judges, breeders/handlers, or people with a sustained interest in dogs). These participants were shown photographs of both human faces and dogs (body profiles – see Figure 4.25) and told to look at each photograph and try to remember it. Analysis revealed that whereas all participants recognized upright faces better than inverted faces, dog experts also recognized upright dogs better than inverted dogs. This finding has been interpreted as supporting the notion that the inversion effect is acquired as a result of expertise and is not a 'face-specific' effect.

What changes then in the way we process faces as we acquire this expertise? Diamond and Carey proposed a distinction between first-order and second-order relational properties. *First-order relational properties* refer to the spatial relationships among parts of the face; for example, the eyes are above the nose and the mouth is below the nose. Faces cannot be distinguished according to their first-order relational properties as they all share the same basic arrangement or configuration. However, first-order relational properties help us detect that a visual stimulus is a face – a necessary step before identifying the face. *Second-order relational properties* refer to the differences in this basic configuration. This refers to the differences in the way the features are arranged in relation to each other; for example, wide-set eyes with a low forehead versus



(a) Inspection items that participants were asked to remember



(b) Recognition items: participants were asked to judge which of the stimulus items they had seen previously

FIGURE 4.25 Examples of the dog stimuli used. *Source: Diamond and Carey, 1986, p.112*

narrow-set eyes and a high forehead. Expertise results in a greater sensitivity to these second-order relational properties, as it is these properties that individuate members of the same class, such as human faces.

There is support for the notion that inversion influences our sensitivity to second-order relational properties. For example, Searcy and Bartlett (1996) presented participants with photographs of grotesque-looking faces. They created images where they had either distorted individual facial features (eyes and mouths) or they had distorted the spatial relations between the features (see Figure 4.26). They then presented these manipulated images in upright and inverted orientations. Participants rated the grotesqueness of the images and results showed that images of faces with distortions to the spatial relations between the features were rated as less grotesque when presented inverted rather than upright; inversion failed to reduce ratings of grotesqueness when the distortions were performed on the features. These findings support the notion that inversion disrupts our processing of spatial relationships between the features.

Research like this suggests that our expertise in (upright) face recognition stems from the way in which these upright faces are processed as 'configurations' rather than as an assemblage of independent features. The term configural processing has been used, although this has been interpreted in a number of ways: to refer to the spatial relationships between features (i.e. second-order relational properties); to refer to the way facial features interact with one another (i.e. the way the shape of the mouth influences how the shape of the nose is perceived); to refer to holistic processing of the face (i.e. the face is perceived

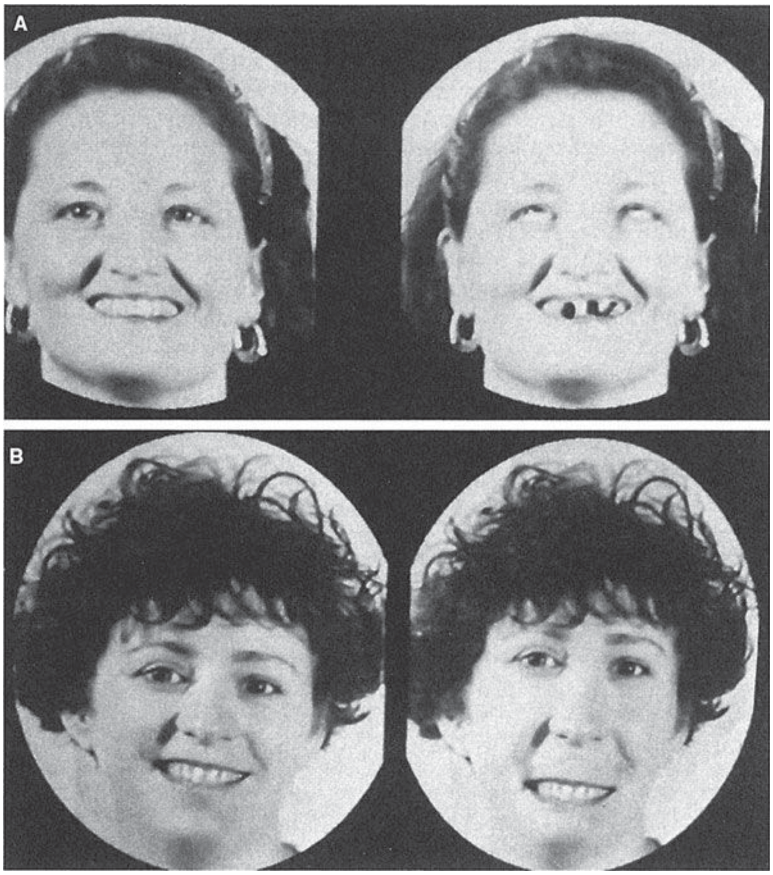


FIGURE 4.26 Examples of stimuli used by Searcy and Bartlett (1996): the pair labelled 'A' shows a normal image and one with distorted facial features; the pair labelled 'B' shows a normal image and one with spatial distortion. *Source: Searcy and Bartlett, 1986, Figure 1, p.907*

as a whole face pattern and not broken down into separate features); or even to refer to the basic arrangement of the facial features (i.e. first-order relational properties). A considerable amount of research has been devoted to investigating the relative importance of this type of processing as compared with the processing of the facial features (known as featural processing or piecemeal processing). Although it is not always clear what different researchers mean by the term 'configural', there is agreement that configural information plays an important role in the perception and representation of upright faces.

One area of debate centres on the suggestion that this reliance on configural processing is the result of learning to recognize lots of faces, and hence the result of expertise. Although this does not rule out input from an innate mechanism, which may have 'kick-started' this learning by biasing attention towards faces, it does not suggest that face perception and recognition involve *unique* processes that are qualitatively different from those used to process other types of stimuli. This generic expertise hypothesis has been criticized by Robbins and McKone (2007), who failed to replicate the finding of Diamond and Carey (1986). Robbins and McKone tested dog experts who they confirmed had good individuation of exemplars of their breed-of-expertise. Their experts included judges, breeders, and trainers of Labradors and so their stimuli were of Labrador Retriever dogs. They compared the performance of these experts with novices on a number of tasks, including inversion, and in contrast to Diamond and Carey found no difference between experts and novices for the dog-inversion effect. Their paper not only presented new data but also provided a review of behavioural experiments often used as support for the expertise hypothesis, and they concluded that configural/holistic processing does not occur for objects of expertise and is therefore unique to face processing. They claimed that the findings of investigations looking at configural processing support the idea of domain-specificity for faces; in other words that faces *are* special.

In sum, there is physiological and neuropsychological evidence supporting the existence of areas specialized for processing faces, and evidence suggesting an innate ability to pay attention to faces. Whether or not the processes involved in face recognition are unique is a central question under investigation.

SUMMARY OF SECTION 7

- Neuropsychological and physiological evidence suggests that there are specific areas of the brain that mediate face processing.
- Research on newborn babies suggests an innate ability to attend to faces.
- The inversion effect appears to be linked to our expertise in processing upright faces using configural information.
- We may develop expertise at distinguishing members of other categories of visual stimuli that also involves configural processing.
- Most evidence does not suggest that the processes involved in the perception and recognition of faces are unique.

8 CONCLUSION

In this chapter we have explored different types of recognition and looked at some of the mechanisms that allow us to recognize objects and faces. In reading about recognition, you may well have got the idea that cognitive psychologists still have a lot to learn about how object and face recognition may occur. Indeed, research has yet to clarify the different processes involved in recognizing familiar faces as opposed to unfamiliar faces, or to specify fully the overlap, if any, between the processes involved in face identification and those used in object recognition.

It is undoubtedly the case that a great deal of research is still being conducted in order to provide a more comprehensive and detailed theory of the cognition involved in recognition. Just as there are different types of recognition, there are also different ways of recognizing faces and objects – for example, visually or by touch – and these different ways may involve different processes. So, rather than seeing the theories discussed here as providing a final answer, the best way to view them is as taking some of the initial steps in this complex but interesting field.

FURTHER READING

Bruce, V., Green, P.R., and Georgeson, M.A. (2003) Visual Perception: Physiology, Psychology and Ecology, Hove, Psychology Press. This textbook on visual perception by Bruce, Green, and Georgeson covers both perception and recognition and provides a more extensive look at the areas of cognition covered in both this and the preceding chapter.

Marr, D. (1982) Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, San Francisco, W.H. Freeman & Company. The critical importance of the work of Marr for theories of perception and recognition is hopefully clear by now. Marr's seminal book, published in 1982, is still widely available and the best way of really getting to grips with his research.

Hole, G. and Bourne, V. (2010) Face Processing: Psychological, Neuropsychological, and Applied Perspectives, Oxford, Oxford University Press. This book by Hole and Bourne is an excellent text suitable for the student reader, that provides a comprehensive review of different areas of research on face processing. As well as covering empirical findings from laboratory based research, evidence from neuropsychological investigations is described and relevant applied research is discussed. It also covers theories of face perception and recognition that we have not had time to go into here, most notably Valentine's multi-dimensional face space model and the approach to face recognition based on principal component analysis.

REFERENCES

Bahrick, H.P. (1984) 'Memory for people', in Harris, J.E. and Morris, P.E. (eds.) *Everyday Memory, Actions and Absent-mindedness*, Academic Press, London.

Bahrick, H.P., Bahrick, P.O., and Wittlinger, R.P. (1975) 'Fifty years of memory for names and faces: a cross-sectional approach', *Journal of Experimental Psychology: General*, vol.104, pp.54–75.

Bauer, R.M. (1984) 'Autonomic recognition of names and faces in prosopagnosia: a neuropsychological application of the guilty knowledge test', *Neuropsychologia*, vol.22, pp.457–69.

Biederman, I. (1987a) 'Recognition by components: a theory of human image understanding', *Psychological Review*, vol.94, pp.115–47.

Biederman, I. (1987b) 'Matching image edges to object memory', *Proceedings of the First International Conference on Computer Vision*, IEEE Computer Society, pp.384–92.

Biederman, I. and Cooper, E.E. (2009) 'Translational and reflectional priming invariance: a retrospective', *Perception*, vol.38, pp.809–825.

Biederman, I. and Gerhardstein, P.C. (1993) 'Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance', *Journal of Experimental Psychology: Human Perception and Performance*, vol.19, pp.1162–82.

Biederman, I., Yue, X., and Davidoff, J. (2009) 'Representation of shape in individuals from a culture with minimal exposure to regular, simple artifacts: sensitivity to nonaccidental vs. metric properties', *Psychological Science*, vol.20, pp.1437–42.

Bruce, V. (1982) 'Changing faces: visual and non-visual coding processes in face recognition', *British Journal of Psychology*, vol.73, pp.105–16.

Bruce, V. and Young, A. (1986) 'Understanding face recognition', *British Journal of Psychology*, vol.77, pp.305–27.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P.J.B., Burton, A.M., and Miller, P. (1999) 'Verification of face identities from images captured on video', *Journal of Experimental Psychology*: *Applied*, vol.5, pp.339–60.

Bulthoff, H.H. and Edelman, S. (1992) 'Psychophysical support for a two-dimensional view interpolation theory of object recognition', *Proceedings of the National Academy of Sciences of the USA*, vol.89, pp.60–4.

Burton, A.M. and Bruce, V. (1993) 'Naming faces and naming names: exploring an interactive activation model of person recognition', *Memory*, vol.1, pp.457–80.

Burton, A.M., Bruce, V., and Johnston, R.A. (1990) 'Understanding face recognition with an interactive activation model', *British Journal of Psychology*, vol.81, pp.361–80.

Burton, A.M., Miller, P., Bruce, V., Hancock, P.J.B., and Henderson, Z. (2001) 'Human and automatic face recognition: a comparison across image formats', *Vision Research*, vol.41, pp.3185–95.

Diamond, B.J., Valentine, T., Mayes, A.R., and Sandel, M.E. (1994) 'Evidence of covert recognition in a prosopagnosic patient', *Cortex*, vol.28, pp.77–95.

Diamond, R. and Carey, S. (1986) 'Why faces are and are not special: an effect of expertise', *Journal of Experimental Psychology: General*, vol.115, pp.107–17.

Ellis, H.D. and Lewis, M.B. (2001) 'Capgras delusion: a window on face recognition', *Trends in Cognitive Sciences*, vol.5, no.4, pp.149–56.

Ellis, H.D. and Young, A.W. (1990) 'Accounting for delusional misidentifications', *British Journal of Psychiatry*, vol.157, pp.239–48.

Gibson, J.J. (1986) *The Ecological Approach to Visual Perception*, Hillsdale, NJ, Lawrence Erlbaum Associates.

Hay, D.C. and Young, A.W. (1982) 'The human face', in Ellis, A.W. (ed.) *Normality and Pathology in Cognitive Functions*, London, Academic Press.

Hay, D.C., Young, A.W., and Ellis, A.W. (1991) 'Routes through the face recognition system', *Quarterly Journal of Experimental Psychology*, vol.43A, pp.761–91.

Hirstein, W. and Ramachandran, V.S. (1997) 'Capgras syndrome: a novel probe for understanding the neural representation of identity and familiarity of persons', *Proceedings of the Royal Society, London, Series B*, vol.264, pp.437–44.

Hole, G. and Bourne, V. (2010) *Face Processing: Psychological, Neuropsychological, and Applied Perspectives*, Oxford, Oxford University Press.

Humphreys, G.W. and Bruce, V. (1989) *Visual Cognition: Computational, Experimental and Neuropsychological Perspectives*, Hove, Lawrence Erlbaum Associates Ltd.

Humphreys, G.W. and Riddoch, M.J. (1984) 'Routes to object constancy: implications from neurological impairments of object constancy', *Quarterly Journal of Experimental Psychology*, vol.36A, pp.385–415.

Johnson, M.H. and Morton, J. (1991) *Biology and Cognitive Development: The Case of Face Recognition*, Oxford, Blackwell.

Johnston, A., Hill, H., and Carmen, N. (1992) 'Recognizing faces: effects of lighting direction, inversion and brightness reversal', *Perception*, vol.21, pp.365–75.

Johnston, R.A. and Bruce, V. (1990) 'Lost properties? Retrieval differences between name codes and semantic codes for familiar people', *Psychological Research*, vol.52, pp.62–7.

Kemp, R., Pike, G., and Brace, N. (2001) 'Video-based identification procedures: combining best practice and practical requirements when designing identification systems', *Psychology, Public Policy and Law*, vol.7, no.4, pp.802–7.

Kemp, R., Towell, N., and Pike, G. (1997) 'When seeing should not be believing: photographs, credit cards and fraud', *Applied Cognitive Psychology*, vol.11, no.3, pp.211–22.

Lawson, R. and Humphreys, G.W. (1996) 'View-specificity in object processing: evidence from picture matching', *Journal of Experimental Psychology: Human Perception and Performance*, vol.22, pp.395–416.

Lederman, S.J. and Klatzky, R.L. (1987) 'Hand movements: a window into haptic object recognition', *Cognitive Psychology*, vol.19, pp.342–8.

Lederman, S.J. and Klatzky, R.L. (1990) 'Haptic classification of common objects: knowledge-driven exploration', *Cognitive Psychology*, vol.22, pp.421–59.

Lederman, S.J., Klatzky, R.L., and Pawluk, D.T. (1993) 'Lessons from the study of biological touch for robot haptic sensing', in Nichols, H. (ed.) 'Advanced tactile sensing for robotics', in *World Scientific Series in Robotics and Automated Systems*, vol.5, Singapore, World Scientific Publishing.

Longmore, C.A., Liu, C.H., and Young, A.W. (2008) 'Learning faces from photographs', *Journal of Experimental Psychology: Human Perception and Performance*, vol.34, pp.77–100.

Marr, D. (1977) 'Analysis of occluding contour', *Proceedings of the Royal Society of London, Series B*, vol.197, pp.441–75.

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, W.H. Freeman and Company.

Marr, D. and Nishihara, H.K. (1978) 'Representation and recognition of the spatial organization of three-dimensional shapes', *Proceedings of the Royal Society of London, Series B*, vol.211, pp.151–80.

Megreya, A.M. and Burton, A.M. (2006) 'Unfamiliar faces are not faces: evidence from a matching task', *Memory & Cognition*, vol.34, pp.865–76.

Megreya, A.M. and Burton, A.M. (2007) 'Hits and false positives in face matching: a familiarity based dissociated', *Perception & Psychophysics*, vol.69, pp.1175–84.

Memon, A. and Bruce, V. (1985) 'Context effects in episodic studies of verbal and facial memory: a review', *Current Psychological Research & Reviews*, pp.349–69.

Morrison, D.J., Bruce, V., and Burton, A.M. (2001) 'Understanding provoked overt recognition in prosopagnosia', *Visual Cognition*, vol.8, pp.47–65.

Neisser, U. (1967) *Cognitive Psychology*, New York, Appleton-Century-Crofts.

Pike, G., Brace, N., and Kynan, S. (2001) *The Visual Identification of Suspects: Procedures and Practice: A Publication of the Policing and Reducing Crime Unit*, London, Home Office Research, Development and Statistics Directorate.

Pike, G., Kemp, R., and Brace, N. (2000) 'The psychology of human face recognition', *IEE Electronics and Communications: Visual Biometrics*, 00/018, pp.12/1–12/6.

Robbins, R. and McKone, E. (2007) 'No face-like processing for objects-of-expertise in three behavioural tasks', *Cognition*, vol.103, pp.34–79.

Searcy, J.H. and Bartlett, J.C. (1996) 'Inversion and processing component and spatial-relational information in faces', *Journal of Experimental Psychology: Human Perception and Performance*, vol.22, pp.904–15.

Selfridge, O.G. (1959) 'Pandemonium: a paradigm for learning', in *The Mechanisation of Thought Processes*, London, HMSO.

Sergent, J. and Poncet, M. (1990) 'From covert to overt recognition of faces in a prosopagnosic patient', *Brain*, vol.113, pp.989–1004.

Tanaka, J.W. (2001) 'The entry point of face recognition: evidence for face expertise', *Journal of Experimental Psychology: General*, vol.130, pp.534–43.

Tarr, M.J. (1995) 'Rotating objects to recognize them: a case study on the role of viewpoint dependency in the recognition of three-dimensional objects', *Psychonomic Bulletin and Review*, vol.2, pp.55–82.

Warrington, E.K. and Taylor, A.M. (1978) 'Two categorical stages of object recognition', *Perception*, vol.7, pp.695–705.

Yin, R.K. (1969) 'Looking at upside down faces', *Journal of Experimental Psychology*, vol.81, pp.141–5.

Young, A.W., Hay, D.C., and Ellis, A.W. (1985) 'The faces that launched a thousand slips: everyday difficulties and errors in recognizing people', *British Journal of Psychology*, vol.76, pp.495–523.

Young, A.W., Newcombe, F., De Haan, E.H.F., Small, M., and Hay, D.C. (1993) 'Face perception after brain injury', *Brain*, vol.116, pp.941–59.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

PART 2

CONCEPTS AND LANGUAGE

Introduction

5 Concepts

Nick Braisby

6 Language processing

Gareth Gaskell and Helen Brown

7 Language in action

Simon Garrod and Tony Sanford

INTRODUCTION

In Chapter 1, we saw how cognitive psychology seeks to explain cognition in terms of information processing by developing and refining accounts that are expressed in terms of representations, which carry information, and computations, which transform the representations in various ways. Whereas Part 1 showed how such accounts could be developed to explain perceptual processes, in Part 2 we shall see how successfully this approach can be applied to two related areas of cognition: categorization and language.

Categorization, our ability to group things together into discrete categories such as fruit, vegetables, tables, and chairs, can be examined in different ways. It can be analysed from a perceptual point of view – how can particular visual or auditory features, for example, influence how we categorize the scenes that we perceive? – but also from a linguistic viewpoint – how is our categorization influenced by the information we receive via language and also by the words we have available? In placing categorization in Part 2, we have chosen to emphasize the link between categorization and language, but in making concepts the topic of the first chapter after Part 1 we also hope to draw attention to some of the links between categorization and perception. Indeed, categorization or semantic classification can be seen as the next stage on from perceptual classification, the focus of Chapter 4.

In Chapter 5, Nick Braisby outlines several different theoretical approaches to categorization. Despite being a fundamental ability, categorization appears to elude a comprehensive treatment. The first two theories outlined, classical and prototype theories, imply that concepts, our mental representations of categories, can be neatly demarcated one from another and each understood in terms of sets of features. According to both theories we place items into a category if they possess a criterial number of these features. Such theories are knowledge-lean – that is, they assume first that it is possible to demarcate category-relevant knowledge, and second that only this knowledge is relevant to determining categorization.

However, both of these theories suffer a number of problems. The alternative theories discussed in the chapter assume that categorization is knowledge-rich, that is, it involves broader knowledge structures – lay theories about domains are implicated by the 'theory'-theory of concepts, and beliefs about what constitute essential properties are implied by psychological essentialism.

As broader knowledge structures get invoked to explain categorization, however, you will see that it becomes harder to state theories precisely, and the discussion of theories in the chapter reflects this. Whereas classical and prototype theories are outlined with some precision, so that one can imagine detailed accounts of the process of categorization being given, 'theory'-based and essentialist theories are hard to define, and it is unclear whether a detailed information processing account could be developed at present.

Because of the difficulty in developing precise accounts of representations of categories and the processes constituting categorization, researchers have been led to revisit some of the simplifying assumptions previously made in this literature. Perhaps, for example, there might be different kinds of categorization for different kinds of category, or for different kinds of categorizer. In some sense, researchers are considering again what categorization really is. In the terminology of Marr's levels that we saw in Chapter 1, in spite of its fundamental importance, researchers are still seeking agreement over a level 1 account of categorization – over what categorization achieves and why. Only then might we hope to develop precise algorithmic, or level 2, accounts.

Gareth Gaskell and Helen Brown's Chapter 6 builds on some of the foundations of Chapter 5, but seeks to explain something that superficially appears very different – how we comprehend both spoken and written language. Together with Simon Garrod and Anthony Sanford's Chapter 7, these chapters on language mark a point of departure. The chapters in Part 1, and Chapter 5 to some extent, have been concerned with how we perceive and pick up information concerning our environment, and how we use this information to infer the presence and nature of objects, and the categories to which they belong. Chapters 6 and 7 mark a concern with the social world, with how we communicate about our world to others, how we make sense of the interpretations of others, and how others influence our communication. As Gareth Gaskell states in opening his chapter, understanding our ability to use language is key to understanding what differentiates humans from other animals, and to understanding human cognition.

Chapter 6 draws our attention to many aspects of language processing that we normally take for granted. In comprehending spoken language, we have to infer which words are present in a stream of speech, an ability we learn as children. We also have to learn to make use of our knowledge of the speech sounds used in our particular language(s). These processes are easily taken for granted, and researchers have had to coin new terms and posit new theoretical structures, such as the mental lexicon, in order to make sense of the comprehension process. Researchers have assumed that different kinds of knowledge are brought to bear at different stages of comprehension. Models of the process that incorporate new theoretical structures and different kinds of knowledge have been constructed (e.g. the cohort model) and experiments conducted to evaluate them. Indeed, and in contrast with Chapter 5, Chapter 6 focuses mainly on processing accounts and how well they explain experimental data. Also in contrast with Chapter 5, some of the processing accounts are sufficiently well specified that they have been developed as computational models. TRACE, for example, is a connectionist model, as is IAC (interactive activation and competition), a model you encountered in Chapter 4. That researchers have been able to develop such models successfully is a testament to how advanced their understanding of the cognitive processes of language comprehension is.

Nevertheless, running through Chapter 6, and in common with Chapter 5, is a concern with the extent to which we require general knowledge for processing language, and the time-point at which this knowledge is brought to bear. The bottom-up (and autonomous) view is that semantic knowledge, for example, is only called upon late on in processing, and only then to adjudicate between interpretations of the incoming input. The top-down (and interactive) view is that such knowledge may operate early, and influence which interpretations are pursued. This important debate, ranging over phenomena such as spoken and written word recognition, ambiguity resolution, and sentence comprehension, is as important as it is unresolved.

This debate is also reflected in Chapter 7, where Simon Garrod and Anthony Sanford broaden the focus on language to include the comprehension of whole texts (not just sentences), language production, and dialogue. The authors begin by considering some difficulties for a simple view of language processing in which information is processed at one discrete stage, and then passed to another, and to another. In the simple view, each stage involves interpreting or translating the output of the previous stage. So, for example, in speech comprehension, a phonological stage might be followed by a lexical stage, which might be followed by a syntactic stage, and so on. Each stage takes as its input the output of the previous stage, transforms that input in a certain way, and then outputs to a subsequent stage. The theme that runs throughout Chapter 7 is that understanding and producing language involves much more than the simple view assumes – in particular, the authors show how language processing as a whole involves drawing heavily on general, non-linguistic knowledge. This is a significant contrast with Chapter 6, where it was assumed that knowledge could be neatly compartmentalized, and that language comprehension required the use of only particular kinds of knowledge and only at particular stages.

Simon Garrod and Anthony Sanford begin by showing how the comprehension of texts does not just rely on consulting the meaning of words in the mental lexicon and combining these according to linguistic rules. In various ways, texts require the application of much more than just lexical knowledge – for example, the authors suggest that the meaning of some words is unlikely to be represented in a 'lexicon' but rather is rooted in actual bodily posture and movement. That language can involve more complicated processes is also demonstrated by a discussion of language production. Though there are some similarities to the discussions in Chapter 6 – production can be seen as involving the reverse of comprehension processes – there are also further complexities. Production involves monitoring one's productions to ensure they are as intended, and also designing one's productions according to the social context. And it is the social aspect of language that most clearly comes to the fore in the discussion of dialogue. Dialogue involves coordination between speakers at a number of levels: for example, in taking turns to ask and answer questions, in developing a common understanding, and in what the authors describe as alignment and routinization of representations.

What do these three chapters reveal about the cognitive approach? Perhaps most notable in these chapters is the breadth of the cognitive approach. Researchers tackle very diverse questions – from what knowledge we have of categories to what processes underpin dialogue – but do so from a common perspective – that of seeking to posit mental representations, which carry information, and computational processes that transform them.

The chapters also invite us to think about the success of the cognitive approach. Chapter 6 shows how cognitive psychology has been successful in generating detailed processing models of language comprehension. Chapters 5 and 7 show a different kind of success – though researchers have yet to answer some of the basic questions about categorization and language-in-action, the cognitive approach has helped them to generate different theoretical frameworks and empirical means of examining them. That is, the success of the approach can be measured not only in terms of the success of proposed models, but also in terms of the generation of new research questions.

Finally, the three chapters in this part reveal a reciprocity between the precision with which theories and models may be specified and the extent to which a cognitive process appears to be knowledge-rich; the extent to which it seems to draw on general knowledge. The more general knowledge a process draws upon, the harder it is for researchers to develop precise models. It appears that precision – one of the hallmarks of a scientific account – can be achieved only when the knowledge that influences a cognitive process can be isolated or separated from other kinds of knowledge and demarcated in distinct processing modules. The question as to the modularity of cognition, explicitly addressed in Chapter 20 by Martin Davies and the late Tony Stone, is one to which we shall return again and again.

CHAPTER 5

CONCEPTS

Nick Braisby

1 INTRODUCTION

In the UK some years ago a television programme required people to guess the identity of unusual antique artefacts. Contestants were allowed to hold the objects and discuss their ideas as to what the objects might be, but the objects were chosen so that it was not at all obvious what their function was, nor what they were called. In the parlance of cognitive psychology, they were difficult to **categorize** – objects for which contestants could not readily find an appropriate **concept**.

You can get a sense of the difficulty faced by these contestants from Figure 5.1. This shows some obscure kitchen utensils. Try to guess what they are – answers are given at the end of the chapter.

Normally we categorize things effortlessly. Try looking around you now. Can you see a single object that you can't label or categorize? My guess is that you won't have to think hard to identify the appropriate concept. But how do we achieve this? For, as Figure 5.1 shows, as did the television programme, categorizing something, finding the right concept, can be difficult. In fact, as we shall see, even effortless categorizations are ultimately difficult to explain. The first step is to work out what concepts are.

1.1 Concepts, categories, and words

Dictionaries say that the word 'concept' has different senses. There is a non-technical sense, one that relates loosely to ideas and thoughts. So, for example, we

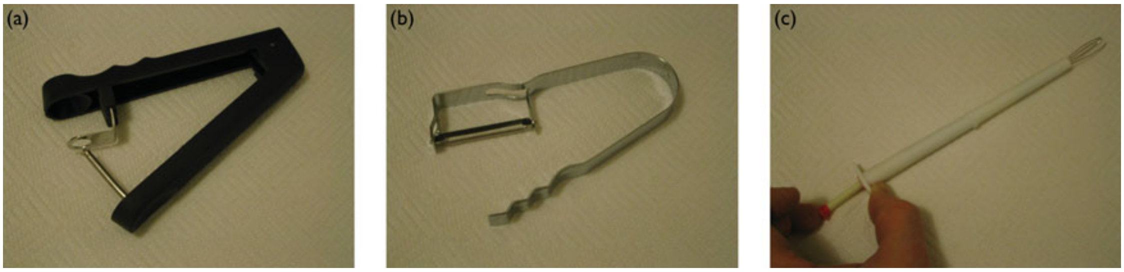


FIGURE 5.1 Three (more or less) obscure objects that may be found in the kitchen.

FIGURE 5.2 Rosie.

might say that a manager has created a new marketing 'concept', meaning he or she has introduced a new idea for promoting a product. However, it is the psychological or philosophical sense that is of interest here. According to this, **concepts** are general ideas formed in the mind: 'general' meaning that concepts apply to every one of a class of things (usually described as a **category**). For example, my concept of 'cat' must be a general idea of all cats – an idea of what cats are in general – not just my pet cat Rosie sitting on the garden fence.

Already this raises two important issues. First, concepts are related to **categories**. Talk of concepts normally presupposes the existence of a corresponding category. There are similarities here with the discussion of Brentano in Chapter 1. Brentano argued that a mental state has two components – a mental act, internal to the mind, and a mental content (the thing that the mental act is about) that is external to the mind. Concepts also have this dual aspect. In thinking 'Rosie is a cat' I perform an activity (thinking) and my thought has a content that is external to the mind – the thought is about Rosie and her relation to the category of (domestic) cats. So, although concepts are internal to the mind, the categories that concepts are about are external. Indeed, researchers standardly adopt the terminological distinction that the word 'concept' refers to something in the mind and 'category' refers to those things in the world that a concept is about.

Second, concepts and categories are linked to words. I used words to communicate the idea that thoughts (such as Rosie is a cat) contain concepts, and that concepts are about categories. Words like 'cat' help you to work out what concept I have in mind (the concept 'cat'). However, it would be too simplistic to suggest that words always pick out concepts straightforwardly. Ambiguous words link to more than one concept – 'chest' relates both to the concept of a body part (torso) and to the concept of furniture (as in chest of drawers). Most words are polysemous – they have many distinct but closely related senses. 'Cat' can refer to the category of domestic cats, but also to big cats and to all felines. Concepts, unlike words, do not have multiple senses, since they are general ideas about *particular* categories. So, we probably have several concepts that all link to the word 'cat' – a concept of 'domestic cat', a concept of 'feline', and so on. Mapping the precise relationships between concepts and words is not at all easy, so for much of this chapter I will assume, as most researchers do for practical reasons, that words pick out concepts in a straightforward manner. Towards the end of the chapter, though, I will try to show some of the complexity of this relationship.

Having considered some preliminaries, we can now turn to the kind of evidence psychologists have used to infer the nature of concepts.

1.2 Categorization

Bruner *et al.* (1956) suggested that 'to categorize is to render discriminably different things equivalent, to group objects and events and people around us into classes, and to respond to them in terms of their class membership rather than their uniqueness' (p.1). According to this definition, concepts are at work whenever people show similarities in behaviour toward different objects in the same category and whenever they show differences in behaviour toward objects in different categories. If, for example, you pat two different dogs, you behave similarly towards them, in spite of their differences. On these definitions you do so because you treat them as instances of the category 'dog'. Likewise, patting a dog but not a house plant signals that you treat these as members of different categories.

Even though concepts may be at work almost all of the time, our focus will be on a restricted range of behaviours that involve giving fairly explicit and often linguistic judgements about category membership. This kind of categorization behaviour has provided the primary evidence as to the nature of concepts.

Categorization could be more broadly construed, however. Potter and Wetherell (1987) and Edwards and Potter (1992) show how attention to natural discourse reveals many subtleties in how people choose which category words to use, and how they then use them in particular contexts. This 'discursive' approach can show how categorization is affected by social influences, such as the social status of the people discoursing, and how using category words serves broader goals than merely that of reporting one's beliefs about category membership. Though this line of work reveals important dimensions to our categorization behaviour, it would be wrong to overplay the significance of such influences. Indeed, it is precisely because categorization behaviour is systematic that people have posited the existence of stable concepts. The focus for cognitive psychologists, therefore, is on categorization processes in general, i.e. what might be common to different instances of categorization in different contexts.

Categorization also need not be so closely tied to language. Indeed, many researchers attribute concepts to non-linguistic animals. Sappington and Goldman (1994) investigated the abilities of Arabian horses to learn to discriminate patterns. They claimed that horses that learned to discriminate triangles from other shapes had actually acquired a concept – in this case, the concept of 'triangularity' – as opposed to merely having learned the particular triangular patterns to which they had been exposed. Again, though, this chapter focuses on what we know of human cognitive achievements, and hence the nature of human concepts.

My bracketing off of these two issues does not solely reflect a pragmatic desire to get on with discussing the matters of most relevance to cognitive psychology, but also the contentious nature of these issues. For instance, some cognitive scientists have argued that the idea that animals might possess concepts is not actually coherent (Chater and Heyes, 1994; though see Newen and Bartels, 2007). Similarly, others have argued against a strong discursive position, according to which categories are socially constructed (e.g. Pinker, 1997; Fodor, 1998).

So, accepting that judgements of category membership are the principal indices of categorization, we can now turn to some of the techniques psychologists have used to elicit these. One method is the sorting task (see Coxon, 1999). In this task, participants are shown an array of different items (sometimes words printed on cards) and asked to sort them into groups. Ross and Murphy (1999), for example, used this technique to examine how people categorize foods. They found that people sometimes put eggs in the same group as bacon and cereal (suggesting a category of breakfast foods), whereas at other times they put eggs together with butter and milk (suggesting a category of dairy products). The groups into which items are categorized are taken to reflect corresponding concepts. The fact that eggs are sometimes put into different groups is consistent with Barsalou's (1983) findings that categorization can depend upon people's goals or purposes. So, for example, when asked to say what falls into the category 'things to take with you in case of fire' people would mention items that would not normally be found together in the same category (e.g., loved ones, pets, and family heirlooms).

If the sorting task seems abstract and artificial, go into your kitchen and look at how the different foods and gadgets are organized. You will probably find items grouped into categories – herbs and spices in one place, fruit in another, vegetables in yet another. I group foods together even in my supermarket carrier bags – usually into nothing more complicated than frozen, chilled, and room temperature – when the person at the checkout gives me enough time to do so! So, placing members of a category together is really an everyday activity that the sorting task taps in a more measurable and controlled way.

The sorting task and the property-listing technique in Activity 5.1 highlight a third aspect of concepts – they are invoked to *explain* categorization behaviour. We behave differently with cats and dogs because cats and dogs belong to different categories, and so our concepts of cats and dogs must differ. The differences

✱ ACTIVITY 5.1

Think of the properties that dogs have. You might think dogs 'bark', 'have four legs', 'run after sticks', 'pant', and so on. Now consider the concept of 'cat'. Take two minutes to write down some of the properties cats have. Don't dwell on any particular property; just write down whatever comes to mind. If you get to ten properties, stop.

COMMENT

Simple though this task seems, it gets hard to think of new properties after a while. Psychologists use this property- or attribute-listing technique to investigate people's concepts, although they obtain results from many participants for each category. They then compare the lists from different people and generate a further list of the most frequently mentioned properties. This gives an indication of the information incorporated in people's concepts, and the frequency of mention indicates how central each property is to the concept.

(and similarities) between cats and dogs are reflected in our concepts, and it is these that are involved in producing behaviour.

1.3 The wider story of concepts

Perhaps because concepts are implicated in so much of our behaviour, their role often goes unnoticed. However, concepts are sometimes brought explicitly into focus. Umberto Eco (1999) discusses the example of the platypus. In 1798 a stuffed platypus was sent to the British Museum. Initially, it was considered so strange that it was thought to be a hoax, with its beak artificially grafted onto its body. For the next 80 years the question of how the platypus should be categorized was hotly debated. Finally, in 1884, it was declared to be a type of mammal, called a monotreme, which both lays eggs and suckles its young, and this categorization has stuck (though of course it is possible that even this categorization might again come into question – see Section 2.1.4).

This case of scientific 'discovery' reminds us that all of our concepts have a past. Even such basic concepts as 'human', 'table', and 'food' have a rich, though perhaps not fully discoverable, history. But the example of the platypus shows that categorization can be a complex process. Though everyday categorizations seem effortless and routine, it took the best scientific minds nearly 90 years to decide how the platypus should be categorized.

Box 5.1 offers another example where categorization has been more explicitly discussed; legal and moral cases provide others. In the UK, for example, the law applies differently to adults and children. So, it is important to be able to categorize everyone as either a child or an adult. Yet it is too difficult to identify a precise age for the boundary between children and adults, and so parliaments have to decide, arbitrarily, where it should lie.

Fascinating though these examples are, the rest of the chapter concentrates on more everyday categorizations. Researchers have tended to adopt a methodological strategy of explaining the simpler cases first, in the hope that explanations can then be developed for more complex cases. As you will see, even everyday categorizations are surprisingly difficult to explain.

1.4 Concepts and cognition

In the last chapter you saw that the word 'recognition' labels different kinds of process. The authors focused on what was called **perceptual classification** and you may have wondered about the subsequent stage labelled semantic classification. Well, semantic classification is what concepts are for. So, the use of concepts to classify – for example, using the concept of 'cat' to classify or categorize my pet cat Rosie – can be viewed as a further kind of recognition.

BOX 5.1 Categorization and diagnosis

Clinicians need to categorize conditions and diseases in order to treat their patients. Though we usually call this diagnosis, it is really a form of categorization – clinicians consider the various properties or symptoms that a patient manifests and attempt to categorize or diagnose the underlying condition. For example, diagnosing or categorizing chronic fatigue, or ME, is notoriously difficult. Macintyre (1998) suggests diagnosis should be based on major criteria – chronic unexplained fatigue that is debilitating and that is not due to exertion, nor substantially alleviated by rest. She also suggests that at least four out of eight minor criteria should be present (e.g. sore throat, muscle pain).

Categorization can also be seen in the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 1994), which gives criteria for diagnosing different mental illnesses. According to the manual, a diagnosis of schizophrenia should be made on the basis of characteristic symptoms, social or occupational dysfunction, duration, etc. Although the manual lists five characteristic symptoms (e.g. delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behaviour) it indicates a diagnosis of schizophrenia may be made when only two are present.

You will see later that both of these kinds of diagnosis, which require only a certain number of a longer list of symptoms to be present, relate to a particular theoretical approach to concepts. Though our discussion of concepts is rooted in laboratory-based studies, it is just a short step to matters of practical import.

Concepts can also be seen as the basic units of semantic memory. While episodic memory stores memories of particular episodes, such as what happened on your last birthday, semantic memory is our long-term memory for facts about the world, such as ‘cats are animals’. The episodic-semantic distinction, which you have already encountered in Chapter 2, is discussed in much more detail in Chapter 8. For our purposes, we simply note that elements of semantic memory such as ‘cats are animals’ express relationships between concepts (between ‘cat’ and ‘animal’ in this case).

We have already mentioned the relationship between concepts and words, but many researchers assume a more explicit link. It is thought that some concepts, called **lexical concepts** (i.e. concepts for which there is a single word), represent our understandings of the meanings of words and are stored in something called the mental lexicon (see Chapter 6). For example, our concept of ‘cat’ would represent what we believe the word ‘cat’ means. The process of understanding language therefore partly involves retrieving lexical concepts from the mental lexicon. Of course, this is a complex process; there may be several lexical concepts corresponding to a single word like ‘cat,’ so we would also have to identify which lexical concept is most appropriate. These and other complexities are developed in Chapter 6.

Concepts also play a role in reasoning. Your list from Activity 5.1 indicates some of the information in your concept of ‘cat’. You may have written things like ‘meows’, ‘likes fish’, ‘mammal’, etc. You might not have written ‘has a heart’ but this is a property of cats too. Now suppose someone asked you whether Rosie has a heart. My guess is that you would say she does. But this is curious, because I have told you only that Rosie is a cat. How have you managed to draw the inference that she has a heart? The answer, of course, is that your concept of ‘cat’ indicates that cats are mammals, and your concept of ‘mammal’ indicates that mammals have hearts. From these concepts you can infer that cats, like Rosie, have hearts. Such inferences might not always be valid of course – though I don’t doubt Rosie has a heart, for all I know, maybe, miraculously, she has some complex artificial pump instead. The complexities of reasoning, of drawing inferences, are the topic of Chapter 12.

Because concepts allow us to make inferences, they simplify the task of remembering information. If you want to remember the properties of Rosie, you would do well simply to remember that she is a cat. If she were unusual in some way (such as having a piece of her ear missing) you might have to remember that information too. But you do not need to remember explicitly that Rosie meows, or that she has a heart, because you can draw these inferences simply by knowing she is a cat. Suffice it to say that our ability to store concepts in semantic memory, together with our ability to reason and draw inferences, simplifies the task of remembering information. Here concepts, reasoning, and memory all act together.

SUMMARY OF SECTION 1

- Concepts are ideas in the mind that are about categories in the world.
- Words tend to pick out concepts, though the exact relationship between them is complex.
- The principal evidence for concepts comes from categorization behaviour, which involves people making judgements concerning category membership.
- Concepts play a wide role in cognition.

2 EXPLAINING CATEGORIZATION

How do we decide that some items belong to the same category and other items belong to different categories? What is it about different cats, for example, that makes us think they are all ‘cats’ and not ‘dogs’?

2.1 Similarity I: the classical view of concepts

According to the classical view of concepts, which has its roots in the philosophical writings of Aristotle (Sutcliffe, 1993), things belong to categories because they possess certain properties in common. There are two aspects to this idea. First, if something is a member of a category, then it must possess the properties common to the category’s members. Second, if something possesses the properties common to a category’s members then it too must be a member of the category. The first aspect asserts that possession of the common properties is necessary for category membership; the second indicates that possession of the common properties is sufficient for category membership. The classical view, then, is that there are necessary and sufficient conditions on category membership. Another way of expressing this is to say that the classical view is that concepts provide definitions of their corresponding category.

In this view, categorization is explained in terms of a comparison of any putative instance with the conditions that define the category. If the instance matches the concept on each and every condition, then it falls within the category – it is a member of the category. If it fails to match on any condition, then the instance falls outside the category – it is a non-member. Let’s consider an example – the category of bachelors. The classical view contends that the category can be defined; that there are properties that are both necessary and sufficient for membership. Might this be true? Dictionaries tell us that bachelors are unmarried adult males. Perhaps these properties are necessary and sufficient for bachelorhood. If they are, then any person who is a bachelor must also be unmarried, adult, and male. Conversely, any person who is unmarried, adult, and male must be a bachelor. And this seems right; it doesn’t seem possible to imagine a bachelor who is married, say. Nor does it seem possible to imagine someone who is unmarried, adult, and male who isn’t a bachelor.

✱ ACTIVITY 5.2

Consider the categories sparrow, gold, chair, introvert, red, and even number. Can you provide definitions for them? Take a few minutes to list the properties for each that you think are important for category membership. Don't worry if you find this difficult; just write down what comes to mind. If you can't think of anything, pass on to the next category. When you have finished try to answer the following questions. First, do you think each of the properties is necessary for category membership (i.e. must every member of the category possess the property)? Second, are the properties for each category, when taken together, sufficient for membership in the category (i.e. must anything that possesses these properties necessarily be a member of the category)?

COMMENT

Most people find this kind of activity difficult. In spite of the classical view, it is surprisingly difficult to think of watertight definitions – you might have succeeded for 'even number' but perhaps not for the other categories. We will consider this again later.

The classical view was supported by some early, empirical investigations (e.g. Hull, 1920; Bruner *et al.*, 1956) that showed people categorized instances according to whether they possessed the necessary and sufficient conditions of the category. However, despite being sporadically defended (e.g. Sutcliffe, 1993) there have been numerous criticisms, and the view has generally fallen into disrepute. The first criticism concerns the phenomenon known as typicality.

2.1.1 Typicality effects

Since the classical view contends that all members of a category must satisfy the same definition, it follows that they should all be *equally good* members of that category. However, psychologists have found systematic inequalities between category members. Rosch (1973) elicited participants' ratings of the typicality or 'goodness-of-exemplar' (sometimes referred to as GOE) of particular instances of a category – a method often known as a **typicality ratings** method. Rosch's instructions give a sense of what is involved.

> *Think of dogs. You all have some notion of what a 'real dog', a 'doggy dog' is. To me a retriever or a German Shepherd is a very doggy dog while a Pekinese is a less doggy dog. Notice that this kind of judgement has nothing to do with how well you like the thing . . . You may prefer to own a Pekinese without thinking that it is the breed that best represents what people mean by dogginess. (pp.131–2).*

Rosch's results for four different categories are shown in Table 5.1. She took these ratings to be indicative of the internal structure of categories, and this conclusion was supported by other empirical work. For instance, Rips *et al.* (1973) and Rosch (1975) examined the relationship between typicality and the time it takes participants to verify sentences that express categorization judgements. The method is often known as **category** or **sentence verification**. For example, the sentences might be 'a robin is a bird' (typical instance) and 'a penguin is a bird' (atypical instance). Participants were asked to respond either 'Yes' (meaning they thought the sentence was true) or 'No' (meaning they thought it was false) as quickly as possible. The results showed that the more typical the instance being considered, the quicker people were to verify the sentence (i.e. the sentence 'a robin is a bird' was verified more quickly than the sentence 'a penguin is a bird').

Further support for the idea that categories have internal structure came from Rosch and Mervis (1975). They used a method known as property- or **attribute-listing** – the method you tried in Activity 5.1. They asked their participants to generate lists of properties for a series of category instances, e.g. robin and penguin for the category bird. The results showed that less typical instances shared properties with fewer category members, while more typical instances shared properties with many other instances. For example, robins have properties – flying, eating worms, building nests – that are shared with many other birds.

✱ ACTIVITY 5.3

Now that you have read Rosch's instructions, write down the following words on the left-hand side of a sheet of paper, putting each word on a new line: pineapple, olive, apple, fig, plum, and strawberry. Then, to the right of each word, write down the number (between 1 and 7) that best reflects how well the word fits your idea or image of the category FRUIT. A 1 means the object is a very good example of your idea of what the category is, while a 7 means the object fits very poorly with your idea or image of the category (or is not a member at all).

COMMENT

When you have written down your answers, compare your ratings with those of Rosch shown in the first column of Table 5.1. How might you explain these ratings? Many people feel that their ratings reflect how familiar they are with particular instances, or how frequently those instances are encountered. You might think that in a society where figs were more commonplace than apples, for example, the typicality of these items would be reversed. In a series of studies, Barsalou (1985) investigated the influences of familiarity and frequency on typicality. Contrary to what one might think, he found that typicality did not correlate with familiarity, and only correlated with frequency to a limited extent. So, it seems that even if penguins were much more common than they are and we were all much more familiar with them, we would still think of them as atypical birds!

TABLE 5.1 Rosch's (1973) typicality ratings for various instances of four categories

Fruit		Vegetable		Sport		Vehicle	
Apple	1.3	Carrot	1.1	Football	1.2	Car	1.0
Plum	2.3	Asparagus	1.3	Hockey	1.8	Scooter	2.5
Pineapple	2.3	Celery	1.7	Gymnastics	2.6	Boat	2.7
Strawberry	2.3	Onion	2.7	Wrestling	3.0	Tricycle	3.5
Fig	4.7	Parsley	3.8	Archery	3.9	Skis	5.7
Olive	6.2	Pickle	4.4	Weight-lifting	4.7	Horse	5.9

Penguins have properties – swimming, not flying – that are shared with relatively few other birds.

Using methods such as these, Rosch, Mervis, and others provided impressive evidence that categories have what we might think of as a rich **internal structure**. A definition serves to demarcate members of a category from non-members, but even things inside the category are highly structured. Both penguins and robins would satisfy the definition of a bird, but there are important systematic differences between them that are reflected in the cognitive processes governing categorization. And this seems contrary to the classical view's suggestion that all category members must equally satisfy a category's definition. How can categories have highly typical and atypical members if the classical view is correct? And how strongly does such evidence count against the classical view?

Though the classical view makes strong claims about the membership of categories – membership should be all-or-none – it says nothing about their internal structure. So, the findings of rich internal structure do not show the classical view to be wrong, unless, of course, internal structure reflects category membership. If a penguin were not only a less typical bird than a robin, but also less of a category member than a robin, then ratings of typicality might reflect a graded notion of category membership in which categories have some clear members, some clear non-members, and a range of cases in between. Then, category membership, quite palpably, would not be all-or-none. On the other hand, if typicality does not reflect graded membership, it may be compatible with the classical view. However, typicality effects do expose an inadequacy – even if they do not contradict its basic tenets, it is not at all obvious how the classical view might explain typicality effects. At the very least, the classical view would need supplementing.

2.1.2 Borderline cases

If membership in a category is 'all-or-none' then there should be no **borderline** cases: an item should either satisfy the definition of a category or it should not. Intuition alone tells us there are items whose category membership is unclear. Colour categories, for example, have no obvious boundary. It seems impossible to draw a line on the colour spectrum, say, between red and orange. For where does a red shade fade into orange? Rather, in between these two categories, there seem to be shades that are neither unequivocally red nor unequivocally orange, hence our use of phrases such as 'a reddy-orange'.

McCloskey and Glucksberg (1978) provided evidence that confirmed this intuition for a whole range of categories. They used a method of asking for **categorization judgements**. They asked their participants to respond either 'Yes' or 'No' to questions of category membership (such as 'Is a robin a bird?'). Participants were also asked to rate the same instances for typicality. McCloskey and Glucksberg then considered the level of agreement that participants showed in their categorization judgements, both across individuals and within the same individuals over two times of testing. They found that participants readily agreed on highly typical and atypical items, yet disagreed over time and across individuals for some items of intermediate typicality. For example, people rated 'chair' as a highly typical item of 'furniture', and were consistent amongst themselves and over time in judging a chair to be an item of furniture. Similarly, with highly atypical items such as a ceiling they were consistent in judging this not to be an example of furniture. With items of intermediate typicality, such as bookends, they were much less consistent. Some people judged these to be items of furniture, others did not, and some people changed their judgements across the two times of testing. McCloskey and Glucksberg thus gave empirical weight to the intuition that many categories have borderline cases.

How telling is this evidence? The classical view certainly implies that categories should have no borderline cases. However, it is at least possible that some of the instances that appear borderline are not genuinely indeterminate, unlike the case of colour categories. It might be that patterns of disagreement reveal a lack of knowledge. For example, you may not know whether a tomato is a fruit or a vegetable. Perhaps sometimes you will say it is a fruit, and other times you might say it is a vegetable. But if you consult a dictionary, you will be told that it is a fruit, even though it is usually used as a vegetable (e.g. in sauces). So, it is possible that an instance definitely belongs to one or another category (i.e. *not* a borderline), but uncertainty makes the item appear borderline. Another possibility is that inconsistency reflects perspective-dependence. It might be, for example, that you know that a tomato is technically a fruit, but your categorization judgement is influenced by the fact that it is used mostly as a vegetable. So, you might agree, in a culinary context, that a tomato is a vegetable, but disagree in the context of a biology lesson.

Though these remain logical possibilities, it is not obvious that McCloskey and Glucksberg's examples actually did involve uncertainty or perspective-dependence. Though people disagreed about whether bookends count as furniture, it seems implausible that they did not have enough information or were adopting different perspectives. So, in the absence of alternative explanations, the compelling evidence for borderline cases seems to undermine the classical view.

2.1.3 Intransitivity of categorization

A further source of difficulties for the classical view has been the observation of **intransitivity** in categorization judgements. Transitivity is observed with many relationships: the relation 'taller than' is transitive because if 'A is taller than B' and 'B is taller than C', then it simply follows that 'A is taller than C'. The relationship is 'transitive' because the last statement follows from the first two.

Is categorization transitive? That is, if As are members of category B, and Bs are members of category C, does it follow that As are also members of category C? According to the classical view it does (and perhaps your intuition agrees). The classical view holds that membership in a category is all or none – if an instance falls into a category, it does so unequivocally. So, if rabbits are mammals then, according to the classical view, they possess the defining features of mammals and so are mammals unequivocally. Likewise, if mammals are animals, then they possess the defining features of animals and so are animals unequivocally. There can be no exceptions. So it should just follow, unequivocally, that rabbits must also be animals.

Hampton (1982), however, showed that people's categorization judgements are not in general consistent with transitivity. For example, he found that participants would agree that 'car seats are a kind of chair' and that 'chairs are a kind of furniture' but not agree that 'car seats are a kind of furniture'. Similarly, people might agree that Big Ben is a clock and that clocks are furniture, but not that that Big Ben is an item of furniture. The fact that people strongly reject the transitive inference in these cases represents a real problem for the classical view.

2.1.4 The lack of definitions

Wittgenstein (1953), in developing his account of language-games, considered the idea, as implied by the classical view, that there are common properties to all instances of the category of game.

> *Consider for example the proceedings that we call 'games'. I mean board-games, card-games, ball-games, Olympic games, and so on. What is common to them all? – Don't say: 'There* must *be something common, or they would not be called "games"' – but* look and see *whether there is anything common to all. – For if you look at them you will not see something that is common to* all, *but similarities, relationships, and a whole series of them at that…*
>
> *I can think of no better expression to characterise these similarities than "family resemblances"; for the various resemblances between members of a family: build, features, colour of eyes, gait, temperament, etc. etc. overlap and criss-cross in the same way. – And I shall say: 'games' form a family.*
>
> (Wittgenstein, 1953, pp.66–7)

If Wittgenstein is right, then the classical view is simply mistaken. Whereas it contends that categories have common properties, Wittgenstein's position is that most categories are like 'game' – when you look closely for common properties, you find none. Recall Activity 5.2. There you tried to offer definitions of categories such as 'red' and 'introvert'. Most people find this task difficult, except perhaps for the category 'even number', where there is a rule that defines category membership. Wittgenstein suggests that most categories are really indefinable. Indeed, his position makes sense of a striking anomaly: despite the classical view having a long history, people have identified very few examples of categories that can be defined. Most researchers are forced to fall back on one of a very few examples – my choice of 'bachelor' is particularly hackneyed! I couldn't use another example, such as 'tree', 'river', 'chair', or 'ship', because no one has identified definitions for these categories.

Nonetheless, Wittgenstein has not proved that natural categories cannot be defined, and so it is possible that someone might yet provide definitions. But the philosophers Kripke (1972) and Putnam (1975) undermined even that idea. They considered what would happen if something that was taken to be 'definitional' was later found to be wrong. Consider the concept of 'cat'. Most people would say that cats are mammals, that they have fur and meow, and so on. Are these necessary properties of the category? Well, perhaps there are some cats that don't meow, and some that don't have fur, but surely all cats are mammals – almost by definition one might say. Putnam considered the implications of discovering that all cats are really robots controlled from Mars (i.e. not mammals at all). This is a thought experiment, of course, so don't worry that the scenario is improbable, or even impossible. The critical issue is what would be the implications of such a 'discovery'. In particular, would the things that we had previously called cats still be cats? What do you think? If you had a pet cat, would it still be a 'cat' after this discovery? Kripke and Putnam believe that it would – those things we called cats before the discovery are still cats afterwards (i.e. the robots controlled from Mars are still cats). But since robots aren't mammals, 'being a mammal' could not be a defining feature of cats, even though we previously thought it was! The conclusion that Kripke and Putnam draw is that we might be shown to be wrong about virtually any property that we happen to believe is true (or even defining) of a category.

This is how Pinker puts it: 'What is the definition of *lion*? You might say "a large, ferocious cat that lives in Africa." … Suppose scientists discovered that lions weren't innately ferocious … Suppose it turned out that they were not even cats … you would probably feel that these … were still really lions, even if not a word of the definition survived. Lions just don't *have* definitions.' (Pinker, 1997, p.323, original emphasis).

There are less fanciful examples that convey the same point. People thought the platypus (Section 1.3) bizarrely combined the features of birds (a bill), amphibians (swimming), and mammals (fur). Suppose that some people came to believe, erroneously, that the platypus was really a strange kind of bird. What Kripke and Putnam argue is that in a case like this, no matter how strongly held the belief, it could never be *definitional* for these people that a platypus is a bird. If it were, then as soon as it was determined that the platypus was a mammal after all, by the very same definition it would no longer be a platypus. The arguments of Kripke and Putnam hinge on the intuition that the platypus will still be a 'platypus' no matter what we come to believe, and no matter how wrong those beliefs ultimately turn out to be. If so, then our beliefs about natural categories never really amount to definitions and the classical view must be mistaken.

2.2 Similarity II: prototype theories of concepts

The combined weight of evidence calling into question the classical view led researchers to consider alternatives. Observations of typicality effects suggested to some that concepts are organized around a measure of the central tendency of a category, otherwise known as the **prototype**. Sometimes the prototype may correspond to an actual instance, but in general it is like a 'best' category member, formed by statistically aggregating over examples of the category one encounters. Rosch, for instance, believed that it is a feature of the natural world that certain attributes or properties tend to correlate or cluster together, and it is these natural clusters of correlated attributes that prototypes describe. For example, the prototype for 'bird' might describe the cluster of properties such as having feathers, wings, a beak, and an ability to fly. These properties cluster together in a way that feathers, lips, gills, and an ability to swing through tree branches do not. Whether or not an instance is a category member then depends upon how similar it is to the prototype: an instance falls within the category if it achieves a certain criterion of similarity. If an instance is too dissimilar, if it mismatches on too many properties, then it falls outside the category.

This account is a little like the classical view: both are committed to the idea that similarity explains categorization. For classical theory, instances fall within a category if they match each and every element of the category's definition, and outside the category if they mismatch on any one. The critical difference is that for prototype theories an instance falls within a category even if it mismatches on a number of properties. Though it might not seem dramatic, a simple illustration shows how significant a move this really is. Suppose a category is characterized by five properties (call them A, B, C, D, and E). Now suppose that there is a criterion for membership in the category such that an instance can mismatch on up to, but no more than, two of these five properties. Then there are a number of logical possibilities for category membership, as shown in Table 5.2.

Instance 1 possesses all the characteristic properties of the category. No instance could match on more properties, and so one could think of this as a highly typical, perhaps even a prototypical instance. Instance 2 mismatches on one property and so is less typical. Instances 3 and 4 mismatch on two properties and are

TABLE 5.2 Different kinds of instances (1 to 4) for a category with five characteristic properties. A tick implies that an instance matches on a particular property; a cross implies a mismatch

	Properties				
Instances	A	B	C	D	E
1	✓	✓	✓	✓	✓
2	✓	✓	✓	✓	✗
3	✓	✗	✓	✗	✓
4	✗	✓	✗	✓	✓

less typical again. What Table 5.2 shows is that this category could not be given a simple definition in terms of the five properties; for each property A to E there is an instance of the category that does not possess that property. Hence, not one of A to E is a necessary property. So, although prototype theories could be thought of as having merely relaxed the classical view's criteria for category membership, the upshot is that prototype theory might be able to explain category membership for the many categories that resist definition. (Note the similarities between Table 5.2 and the discussion of diagnosis in Box 5.1 – can you see how the criteria proposed for diagnosing schizophrenia and ME treat these as prototype concepts?)

Prototype theories have been formulated in different ways. One of the earlier formulations was provided by Smith *et al.* (1988), and this captures many of the qualities found in different versions. Table 5.3 gives their illustration of a prototype representation for 'apple'.

TABLE 5.3 Prototype representation for apple following Smith *et al.* (1988)

Diagnosticity	Attribute	Value	Weight
1	COLOUR	red	25
		green	5
		brown	
		. . .	. . .
.5	SHAPE	round	15
		square	
		cylindrical	5
		. . .	. . .
.25	TEXTURE	smooth	25
		rough	5
		bumpy	
		. . .	. . .

Table 5.3 highlights some of the differences between prototype theories and the classical view. First, there are multiple possible values for each attribute, capturing the fact that no one value is necessary for category membership – apples are typically, but not necessarily, red, for example. Second, diagnosticities indicate the extent to which each attribute is important for deciding category membership. Third, the values are weighted and these weights indicate the extent to which each value contributes to typicality; the highest weighted values are those of the prototype. Categorization depends upon achieving a criterion similarity with the representation of the concept, one that depends on matching properties as before, but now diagnosticities and weights enter into the computation as well (though we don't need to go into detail). Much research continues to be devoted to making formal models of the prototype view and testing these against empirical data (e.g. Minda and Smith, 2011).

Prototype theories can readily explain the typicality effects discovered by Rosch and her co-workers.

1. Instances that differ in typicality are assumed to differ in terms of the weighting of values on which they match the concept. For example, in Table 5.3, a difference in typicality between red and brown apples is reflected in a difference in the weighting for red and brown.
2. Sentences such as 'a robin is a bird' are likely to be verified more quickly than 'a penguin is a bird' because, for high typicality instances, the criterial similarity required for verifying the sentence is likely to be achieved after matching just a few properties. This is because most attributes that match will have higher-weighted values, and so any criterion for category membership will be reached quickly. For low typicality instances like penguin, many attributes will mismatch or will have low weighted values, and so more matches will have to be made before the criterion is reached.
3. Typicality is likely to correlate with how widely category members share attributes. This follows

from the fact that the diagnosticities of attributes and weights of values themselves reflect the statistical distribution of those attributes and values. The more widely shared a value is, the greater is its weight. In Table 5.3, for example, 'round shape' receives a high weight indicating that many (many) more apples are round than square. Since high typicality instances tend to match on high weighted values, it follows that they will also possess properties that are widely shared.

However, despite prototype theory being able to accommodate many of the findings that undermined the classical view, difficulties have emerged.

2.2.1 The meaning of typicality effects

Armstrong *et al.* (1983) considered whether typicality effects occur for concepts that appear to be definitional. Their examples of definitional concepts included 'female', 'plane geometric figure', 'odd number', and 'even number' (as in Activity 5.2). Armstrong *et al.* believed that category membership for these concepts is determined not by similarity to a prototype but by a definition: whether a number is even depends on whether dividing it by 2 yields an integer. Curiously, however, they found a range of robust typicality effects (summarized in Table 5.4), implying that even these apparently definitional concepts have an internal structure; these effects were also found using the sentence verification task.

At first glance Armstrong *et al.*'s data could be taken to imply that even concepts such as odd number are not really definitional after all, but organized around a prototype. However, Armstrong *et al.* didn't regard this as a serious possibility. Instead, they argued that the existence of typicality effects should not be taken as conclusive evidence that category membership is determined by similarity to a prototype. They proposed instead a dual-process model, in which concepts possess a 'core' that is used when we judge category membership and a set of identification procedures that we use to identify instances of a category on particular occasions (often rapidly). Armstrong *et al.* suggested that the classical view might explain the concept's core, while prototype theory explains identification procedures. Unfortunately, inasmuch as this proposal involves both theoretical approaches, it appears to inherit some of the problems faced by each.

TABLE 5.4 Typicality ratings for instances of well-defined categories (from Armstrong *et al.*, 1983)

Even number	Typicality rating	Female	Typicality rating
4	1.1	mother	1.7
8	1.5	housewife	2.4
10	1.7	princess	3.0
18	2.6	waitress	3.2
34	3.4	policewoman	3.9
106	3.9	comedienne	4.5
Odd number	**Typicality rating**	**Plane geometry figure**	**Typicality rating**
3	1.6	square	1.3
7	1.9	triangle	1.5
23	2.4	rectangle	1.9
57	2.6	circle	2.1
501	3.5	trapezoid	3.1
447	3.7	ellipse	3.4

2.2.2 The context-sensitivity of typicality effects

Another difficulty for prototype theory is the observation that typicality effects change with context. If, as Rosch thought, prototypes reflect natural correlations or clusters of properties, one would expect the prototype to be stable.

However, Roth and Shoben (1983) showed that typicality effects are changed by linguistic context. For example, their participants rated the typicality of different farm animals with respect to the category 'animal'. Participants were first presented with a context sentence that emphasized a particular activity, e.g. *Bertha enjoyed riding the animal* or *Bertha enjoyed milking the animal*. The context sentence was then followed by a sentence frame such as *The _ quite liked it too*. Participants were asked to rate the typicality of a list of animal words that would complete the sentence frame. Importantly, the list contained words such as horse and cow that fitted well with one context sentence but not with others – though both words were judged to be possible completions of the sentences. Roth and Shoben found that when the context sentence referred to milking, cows were considered to be more typical 'animals' than horses. However, when the context referred to riding, horses were considered more typical 'animals' than cows. (You might notice similarities with the discussion of priming in Chapter 2.)

Medin and Shoben (1988) also found that typicality judgements change with context. They asked their participants to rate various kinds of spoon for typicality in the category spoon. Participants rated metal spoons as more typical than wooden spoons, and small spoons as more typical than large spoons. Therefore, one might expect that small metal spoons would be most typical of all and that large wooden spoons would be least typical, with small wooden and large metal spoons intermediate. However, while Medin and Shoben found that small metal spoons were more typical than large metal spoons, they found that large wooden spoons were more typical than small wooden spoons. So, the contribution to typicality made by the values large and small depended on whether one was thinking about metal spoons or wooden spoons.

Prototype theories cannot easily explain such demonstrations of the instability of typicality. First, the very idea of instability seems to be at odds with Rosch's claim that prototypes correspond to stable clusters of correlated properties that reflect the structure of the natural world. Second, in connection with Table 5.3, Roth and Shoben's results suggest that the weightings of values and/or diagnosticities of attributes are themselves changeable. However, it is unclear what mechanism could be responsible for such changes. Third, Medin and Shoben's results suggest that the contributions to typicality of different properties (e.g. size and material made from) are mutually dependent. Yet the representation in Table 5.3 assumes that the attributes and values are independent of one another.

2.2.3 Complex concepts

As noted in Section 1.4, it is commonplace to assume that concepts express our understandings of the meanings of words. So, the concept 'red' is assumed to express what we understand as the meaning of the word 'red'; the concept 'car' is thought to provide the meaning of the word 'car'. But this immediately raises the question: what kind of concept provides the meaning of the phrase 'red car'?

Researchers have tried to explain the meanings of phrases and larger linguistic units in terms of **complex concepts**, i.e. combinations of lexical concepts. The meaning of the phrase 'red car' would then be explained in terms of the combination of the constituent lexical concepts 'red' and 'car'. How could concepts combine to yield the meaning of such a phrase?

If concepts are structured around prototypes, then perhaps they could combine through combining their prototypes. The difficulty, however, is that no one really knows how this might be done. Though many suggestions have been made, they all appear to fail for one reason or other. For example, one suggestion has been that the prototype for 'red car' is formed from the prototype for 'red' and the prototype for 'car' (the prototypical red car would therefore be a prototypical car that was prototypically red).

Though this seems a sensible suggestion, and appears to give the right interpretation for 'red car', this could not work in general. Following the same reasoning, the prototypical 'pet fish' ought to be a prototypical fish that is also prototypically pet-like – perhaps something like a cuddly salmon. The real prototypical 'pet fish' is of course more like a goldfish – neither a prototypical pet nor a prototypical fish. More problematic still for combining prototypes, the prototypical 'stone lion' ought to be something like a real lion made of stone, i.e. an impossible object. How could the prototypes for 'stone' (perhaps granite or limestone) and 'lion' (a real lion) combine to give the right interpretation (i.e. a stone statue of a lion)? If you feel that these examples are a little whimsical, take a look at newspaper headlines, as these often use phrases with a similar structure. For example, it is not easy to see how the meaning of 'killer firework' could be explained by the combination of the constituent prototypes: a prototypical killer might be a sadistic criminal, or perhaps a virulent disease; a prototypical firework might be a rocket. How would these prototypes combine to yield the required interpretation? Complex concepts continue to present real difficulties for most theories of concepts (see Fodor, 1998).

2.3 Similarity III: exemplar theories

A third theoretical account of categorization also employs the notion of similarity. According to the exemplar view, instead of being summary descriptions of a category – definitions in the case of the classical view, and knowledge structures such as those in Table 5.3 in the case of the prototype view – concepts are instead representations of the individual members (or exemplars) of the category that we have experienced (Medin and Schaffer, 1978). Brooks *et al.* (1991) examined clinicians' ability to diagnose (or categorize) new examples of skin disorders. Those examples that were similar to previously studied cases were categorized more accurately than those that were dissimilar. This effect persisted after a 2-week delay, and was found both in first-year residents in family medicine (recent graduates from medical school) and in general practitioners with an average of 15 years' experience since graduation. It was also found when participants were required to carefully consider all alternative diagnoses. The conclusion that Brooks *et al.* draw is that categorization is influenced both by explicit rules and by similarity to other exemplars. A similar finding was reported by Allen and Brooks (1991) using artificial categories of cartoon pictures of animals. Even though participants were explicitly told the categorization rule, they were more accurate and faster in categorizing items similar to previous exemplars. Allen and Brooks believed reliance on exemplar-level information was likely to happen only in certain circumstances, but Hahn *et al.* (2010) have argued that such effects are more pervasive.

Demonstrations such as these suggest that exemplar-level information is very difficult to ignore in categorization tasks, and that people categorize items through reference to two sources of information: summary, category-level information and information about exemplars. Just as with other similarity-based views, it has proven possible to develop sophisticated models of categorization that have had much success in accounting for empirical data (see, for example, Nosofsky, 2011).

2.4 Commonsense theories

Both classical and prototype theories explain categorization in terms of similarity using quite simple feature sets. But the problems these theories have encountered have led researchers first to question the importance of similarity, and second to propose that categorization involves much larger knowledge structures, called theories (or **commonsense theories** to distinguish them from scientific ones). The approach has become known as the 'theory'-theory of concepts.

However, before we turn to the 'theory'-theory we should note that similarity-based accounts have achieved considerable success and remain popular. Hampton (1998) conveys some sense of this. Using McCloskey and Glucksberg's (1978) data (both typicality ratings and categorization judgements were collected – see Section 2.1), he examined whether the probability of an item being judged a category member could be predicted from its typicality (reflecting its similarity to a prototype).

Focusing on just the borderline cases, Hampton showed that typicality was a very good predictor, explaining somewhere between 46 per cent and 96 per cent of the variance in categorization probability. So, regardless of the difficulties facing similarity-based accounts, similarity (as measured by typicality) seems to be a good indicator of categorization. Nonetheless, Hampton found other predictors of categorization probability (though none was as good a predictor as typicality). These included lack of familiarity; the extent to which an instance was judged 'only technically speaking a member' of a category (e.g. a dolphin is technically speaking a mammal but superficially appears more similar to fish); and the extent to which participants judged an instance was 'technically speaking not a member' (e.g. a bat is technically speaking not a bird despite superficially appearing more similar to birds than to mammals).

That these last two factors were predictors suggests that categorization draws upon deeper, more theoretical knowledge than just similarity alone.

We now turn to some of the reasons why, in spite of these successes, many researchers have become dissatisfied with the notion of similarity.

2.4.1 Problems with similarity

The philosopher Nelson Goodman identified a number of problems with similarity; indeed, he described it as 'a pretender, an impostor, a quack' (Goodman, 1972, p.437). One concern is with whether similarity genuinely helps us to explain categorization. After all, in prototype theories saying that an instance is similar to the prototype means that the two share some properties in common. But note that this further explication removes the notion of similarity: 'is similar to' becomes translated as 'shares properties with'. So, what explains categorization is not similarity *per se* but the sharing of properties.

However, a further problem arises since there is no obvious limit to the number of properties any two objects may share. Murphy and Medin (1985) ask us to consider the similarity of plums and lawnmowers. 'You might say these have little in common, but of course both weigh less than 10,000 kg (and less than 10,001 kg, . . .), both did not exist 10,000,000 years ago (and 10,000,001 years ago, . . .), both cannot hear well, both can be dropped, both take up space, and so on' (p.292). It seems that, depending on what counts as a relevant property, plums and lawnmowers could either be seen as very dissimilar, or very similar. So, for similarity, explicated in terms of shared properties, to provide meaningful explanations of categorization, we need to know what counts as a property. We need some way of declaring 'lack of hearing ability' as irrelevant in comparing plums and lawnmowers, for example. For Murphy and Medin (1985), observations such as these suggest that similarity is shorthand for something else that explains why categories hang together, or cohere.

2.4.2 The role of commonsense theories

In opposition to similarity-based views, Murphy and Medin argued that concepts are explanation-based, that there is some explanatory principle or theory that unites the category. They offer the example of someone at a party who jumps into a swimming pool fully clothed. You might categorize this person as being intoxicated. But a similarity-based view cannot explain this because your concept of 'intoxicated' is unlikely to include the property 'jumps into swimming pools fully clothed'. So how might we explain the categorization? Murphy and Medin argue that categorizing the person as 'intoxicated' plays a role in explaining their behaviour, that is, in explaining why they jumped into the swimming pool.

Might this explanatory basis be found in categorization more generally? If so, then categorizing a robin as a bird ought to provide some kind of explanation of the robin's properties, analogous to the case of the intoxicated swimmer. Such a categorization does appear to provide a (partial) explanation: knowing that a robin is a bird helps explain why it has feathers and a beak. The explanation is partial, since we could go on to ask why birds have beaks and feathers, but it is an explanation nonetheless. After all, were we to discover that a robin is not a bird, we would want to know why it has feathers and a beak. Without the categorization we would be in need of an explanation.

We noted in Section 2.2.2 that similarity-based approaches cannot easily explain the non-independence of attributes. For Murphy and Medin, relationships between attributes are evidence that our concepts are embedded in larger and broader knowledge structures. Sometimes these structures have been labelled commonsense theories, sometimes merely background knowledge. But if such knowledge structures are at work in categorization, why might people have previously concluded that concepts are similarity-based? Murphy and Medin speculate that many categorization judgements become automatized, particularly when members of the same category have relatively consistent perceptual properties. Under these conditions, the role of our underlying theories becomes obscured, and so we may (erroneously) conclude that categorization is determined by similarity. However, even in these cases, when novel instances emerge (such as robot cats) or where there is disagreement (with borderlines perhaps), we turn to our underlying theories.

What evidence is there that categorization is determined by theories as opposed to similarity? Rips (1989) asked his participants to consider triads of objects. Two objects belonged to distinct categories (e.g. a pizza and a quarter) and were chosen so that participants' largest estimate of the size of one category (quarter) was smaller than their smallest estimate of the other (pizza). Rips then asked his participants to consider a third object, telling them only that it was of intermediate size (i.e. larger than the largest estimated size of a quarter and smaller than the smallest estimated size of a pizza). He asked which of the two other categories this third object was more likely to belong to, and which of the two it was most similar to. The two judgements dissociated: that is, participants judged the object more likely to be a pizza, but more similar to a quarter.

Rips also found dissociations when participants considered, for example, a bird-like creature that came (via mutation or maturation) to resemble an insect, and other dissociations between categorization and similarity have been demonstrated (e.g. Rips and Collins, 1993; Roberson *et al.*, 1999). Kroska and Goldstone (1996) showed their participants scenarios that described a putative emotion. Each scenario constituted a set of phrases such that one phrase was central to one emotion and other phrases were characteristic of a different emotion. For example, one scenario included the phrases 'Threat of harm or death', 'Being accepted, belonging', and 'Experiencing highly pleasurable stimuli or sensations'. The first of these phrases was considered central to the emotion category 'fear'. The remaining two phrases were considered characteristic of the emotion category 'joy'. Kroska and Goldstone found their participants tended to categorize this scenario as an instance of fear (i.e. a member of the category 'fear') but they also judged it to be more similar to an instance of joy. That is, judgements of category membership were influenced by properties considered central to a category, while judgements of similarity were influenced by characteristic properties. Again, these findings show that judgements of category membership can dissociate from judgements of similarity.

It seems that there are deeper reasons for people's categorizations – in the quarter example, perhaps they realized that pizzas can, in principle, be any size, whereas their commonsense theories of coins tell them they are produced to a regulation standard.

2.4.3 Difficulties with the theory-theory

The theory-theory has proved an important and useful way of thinking about concepts. It has, for instance, reminded researchers of difficulties with the notion of similarity, and it has proved to be a useful peg on which to hang a range of disparate findings whose common theme is that categorization is influenced by deeper, causal knowledge of categories, as well as by knowledge of their superficial properties.

However, there are a number of difficulties with the theory-theory. Some of the findings taken to support the theory-theory are really demonstrations that similarity does not always explain categorization, and this does not necessarily imply that theories are what is needed. Moreover, it is not clear what is meant by 'theory'. Whereas similarity-based views could be made relatively precise (see Table 5.3 for instance), formalizing theory-theories seems much more difficult. Some researchers have tried to pin down what is meant by a commonsense theory via a comparison with scientific theories (see Gopnik, 1996). However, other researchers believe such a comparison undermines the idea that commonsense theories are theories at all (see Gellatly, 1997). For example, Murphy (2000) argues that the background knowledge that influences concepts is too simplistic and mundane to be likened to a scientific theory. Indeed, he eschews the term 'theory' in favour of the more neutral 'knowledge'.

A further difficulty with the theory-theory is that it is hard to imagine how combining theories could explain complex concepts. Scientific theories are notoriously difficult to combine. Indeed, for decades, theoretical physicists have struggled to combine theories of electricity, magnetism, and gravity into one unified theory. So how can theories be combined so effortlessly in understanding phrases like 'red car' when they are so difficult to combine in general? Even if we talk of combining knowledge rather than

BOX 5.2 Categorization in development

Support for the idea that knowledge of deeper, causal principles is at work in categorization has come from work looking at children's categorization. Keil (1989), for instance, used both discovery and transformation procedures to examine how children weigh appearance and theoretical properties. For example, in a discovery, children might be told of a novel hybrid animal that looked and behaved just like a zebra. However, they would be told also that it had been discovered that this animal had the insides of a horse and was the offspring of two horses. Younger children (around 4 years of age) tended to judge the animal to be a zebra, whereas older children (around 7 years) tended to judge the animal to be a horse. That is, younger children seemed to be influenced more by the superficial characteristics of the animal (e.g. appearance), and older children more by its biologically relevant properties (e.g. lineage).

Similar results were found using a transformation procedure. Children were told of a raccoon that underwent a series of transformations so that it ended up looking and behaving like a skunk. For example, it might have skunk-like stripes dyed on its fur, and have a surgical implant so that it could emit foul-smelling liquid. Again, younger children seemed dominated by appearance-based properties; they judged that the raccoon was now a skunk. The older children, in contrast, judged that the animal was still a raccoon.

Keil has referred to this age-related change in children's categorization as the 'characteristic-to-defining shift', since he thought the younger children were influenced by properties (i.e. appearances) that were only characteristic of the category, while the older children were beginning to deploy something like the beginnings of a biological theory and were paying attention to properties that were more defining. However, as Murphy (2000) points out, it is probably not the case that the younger and older children have qualitatively distinct styles of categorization. It is more likely that the younger children simply do not know enough about biological categories to work out which properties are characteristic and which are defining.

theories, we are still left with the difficult problem of working out which knowledge gets combined and the mechanism by which this is done.

Given these problems, it is ironic that the theory-based view is motivated in part by difficulties with the notion of similarity. Arguably, it has supplanted this with the equally mysterious notion of a 'theory'.

2.5 Psychological essentialism

Psychological essentialism is one attempt at more precisely formulating the view that categorization is influenced by deeper, explanatory principles. Medin (1989) and Medin and Ortony (1989) suggested that people believe and act as though category members have certain essential properties in common. That is, people categorize things according to their beliefs about essential properties. They may also believe that the essential properties constrain a category's more superficial properties. For example, the essential properties of birds might be thought to involve their genetic make-up, properties that would constrain their appearance and behaviour.

Essential properties can be characterized as those properties that if an object did not possess them, it would not be that object. The essential properties of birds are properties that all birds necessarily possess; if something doesn't possess them, then it isn't a bird. Essential properties may seem rather like the defining properties of the classical view. However, there is one critical difference. According to psychological essentialism most people will not know what a category's essential properties are, but will still believe that the category has some. We might speculate as to what the essential properties are – perhaps for biological categories they would be genetic properties – but, in general, our beliefs will be vague and may turn out to be incorrect. So psychological essentialism proposes that people's concepts may contain a 'place-holder' for an essence – and the place-holder may even be empty, reflecting a lack of knowledge as to what the essential properties might be.

Of course not everyone's place-holder need be empty. Indeed, it is usually thought that discovering essential properties is a job for science. A metallurgist or chemist, perhaps, might uncover the essential properties of gold, just as a biologist might for birds. So, experts may have their place-holders partially or completely filled – they may know (or think they know) the essential properties. But these beliefs may turn out to be in error too, so the place-holder is presumably capable of revision. We can illustrate psychological essentialism with the platypus example of Section 2.1.4. Soon after its discovery, lay-people presumably came to believe that the platypus had a certain essence, but had no idea what this might be (their essence 'place-holder' was empty). Experts at the time might have filled their essence place-holder in different ways: some thought the platypus was essentially an amphibian; others that it was a mammal. But the contents of these place-holders changed as more was learned. Finally, the experts settled on the view that the platypus was mammalian, and as lay-people adopted this view they filled out their essence place-holder accordingly.

Psychological essentialism is consistent with much of the evidence supporting the theory-theory. Much evidence supporting psychological essentialism specifically has come from studies of the development of categorization. For instance, Gelman and Wellman (1991) found that 4- and 5-year-old children believe the insides of objects to be more important than their outsides in determining category membership. For example, they asked children whether a dog would still be a dog if its outsides were removed, and also if its insides were removed. Children thought that instances would remain in the category if the outsides were removed, but not if their insides were removed. According to Gelman and Wellman, children are being essentialist since they believe that something internal, something hidden and 'inner', is causally responsible for category membership.

However, psychological essentialism has not gone unchallenged. Malt (1994) examined the concept of water. If people believe H_2O to be the essence of water, then their categorization of liquids as water should be strongly influenced by the proportion of H_2O those liquids contain. However, Malt found that people's categorizations were strongly influenced by the source of the water, its location, and its function. Indeed, pond water was thought to be 'water' but was judged to contain only 78.8 per cent H_2O; tears were judged not to be 'water' but to contain 88.6 per cent H_2O. So the belief in the presence or absence of H_2O was not the only factor in deciding membership in the category 'water'.

In Section 2.1.4 we considered the arguments of the philosophers Kripke (1972) and Putnam (1975). For example, Putnam argued that even if we discovered that all cats are robots controlled from Mars, they would still be cats. What we didn't note there is that they used thought experiments such as this to support essentialism. Braisby *et al.* (1996) subjected these to an empirical test. They asked participants to give categorization judgements in thought experiments such as Putnam's robot cat. In one condition they were told 'You have a female pet cat named Tibby. For many years people have assumed cats to be mammals. However, scientists have recently discovered that they are *all*, in fact, robots controlled from Mars. Upon close examination, you discover that Tibby too is a robot, just as the scientists suggest.' Participants were then asked to indicate whether they thought that a series of statements were true or false. These included statements expressing essentialist intuitions (e.g. 'Tibby is a cat, though we were wrong about her being a mammal') and statements that expressed the contrary intuition (e.g. 'Tibby is not a cat, though she is a robot controlled from Mars'). Only about half of the participants thought that the essentialist statements were true and the contrary ones false. Moreover, many participants seemed to give contradictory judgements: they either judged both statements to be true, or judged both to be false. Braisby *et al.* argued that these findings did not support essentialism, but implied that concepts change their content according to context and perspective (see Braisby and Franks, 1997). Jylkkä *et al.* (2009) empirically re-examined the Braisby *et al.* data and came to a different conclusion from the original authors. They suggested that people's intuitions offer support for essentialism. However, their participants' responses were actually quite varied, and the greater proportion of responses were not consistent with essentialism. One way of interpreting this debate is that both studies suggest essentialism may accurately characterize some people's categorizations, but that an alternative account of concepts is required to explain categorization in general.

Hampton *et al.* (2007) re-examined some of the dissociations reported by Rips (1989) and mentioned in Section 2.4.2. They found that when key manipulations were presented as between-subject rather than within-subject, the proportions of essentialist responses fell, and overall the majority of participants adopted responses inconsistent with essentialism. They also found important individual differences, in that some participants consistently gave essentialist categorizations, whilst others gave more importance to appearance and behaviour. Hampton *et al.* speculate whether such individual differences may reflect more general and stable cognitive styles.

While the argument for essentialism was originally developed to explain natural concepts (e.g. cat, water), researchers have begun to consider whether the theory applies in other domains also. For example, Bloom (1996) proposed that artefacts such as torches are created with a specific purpose in mind, and that the creator's intention constitutes the essence of the category. However, Malt and Sloman (2007) disputed this account, demonstrating that people often describe artefacts according to their function or use. For example, their participants tended to describe a teakettle as a watering can if it was used as a watering can, even though it was created with the intention that it should heat water. (See also Bloom, 2007, for counter-arguments.)

Essentialism has also been applied to social categories. For example, Haslam and Whelan (2008) have proposed that psychological essentialism plays a central role in the categorization of race and ethnicity, gender, mental disorder, personality, and sexual orientation. Haslam *et al.* (2000) developed essentialist belief scales to examine essentialism thinking in relation to categories, and argued that this centred on two key dimensions: the extent to which a category was seen as a natural kind, and the extent to which the category was seen as coherent and homogeneous. Haslam *et al.* (2002) have also discovered links between such dimensions and prejudice, and Haslam and Levy (2006) revealed a similar structure to people's thought concerning homosexuality. However, such studies largely measure people's beliefs about categories and not categorization *per se*. Examining people's categorization of sexual orientation, Braisby and Hodges (2009) failed to find signs of essentialism, with participants categorizing sexual orientation according to described behaviour and giving little weight to putative essential properties such as genetics.

There has also been mixed evidence concerning the role that expert opinion plays in categorization. Malt (1990) presented people with objects that they were told appeared 'halfway' between two categories (e.g. a tree halfway between an oak and a maple) and asked them to indicate how they would solve the dilemma of categorizing the object. She offered her participants three options. They could 'ask an expert', 'call it whichever you want', or indicate that they could 'tell which it is' if they could only think about it long enough. For pairs of natural categories such as 'robin-sparrow' and 'trout-bass', 75 per cent of participants suggested they would ask an expert, whereas for pairs of artefact categories, such as 'boat-ship', 63 per cent of participants suggested it was possible to 'call it whichever you want'. This evidence suggests that people may be psychologically essentialist for natural categories, at least to some degree, because they recognize that experts may be in a better position to judge categorization when lay people cannot. However, the data overall are not conclusive. Braisby (2001) examined the extent to which people modify their categorization judgements for genetically-modified biological categories when told the opinions of experts. For example, his participants might be asked to consider a genetically modified salmon, and were told either that expert biologists had judged that it was a salmon or that they had judged that it was not. He found that only around half of the participants changed their categorization judgements to conform to the judgements of the biologists. Moreover, around a quarter of participants would change their categorization judgements to conform to those of shoppers (i.e. a group presumed not to be expert with respect to the category's essential properties). Braisby argued that only around a quarter of participants were modifying their categorization judgements because of the biologists' expertise with the relevant essential properties, and so the majority of responses did not provide evidence for psychological essentialism. Indeed, participants seemed to base their judgements on non-essential properties such as appearance and function (as well as genetic make-up).

Braisby and Hanlon (2010) also found that the extent to which people would defer to experts in categorization tasks was related to a measure of their cognitive style. Witkin *et al.* (1962) coined the term 'field dependence' to refer to a distinction in cognitive style that has been variously labelled as analytic-holistic. Braisby and Hanlon found that 89 per cent of field dependent participants deferred in their categorization, as opposed to just 23 per cent of field independent participants. Such a result appears to support the suggestion made by Hampton et al. (see earlier in this section).

Lastly, it should be noted that much of the evidence cited in support of psychological essentialism (e.g. Gelman and Wellman, 1991) only indirectly relates to beliefs in essential properties. Gelman and Wellman, for example, found that children thought that removing the outsides from something like a dog did not alter its category membership, but removing its insides did. However, for these data to support essentialism, a further inference is required, one that relates insides to essences. In a similar vein, Strevens (2000) actually argues that the notion of essence or essential properties is not required to explain empirical data such as these. Of course, psychological essentialists have responded to some of these criticisms, so it seems fair to say that the arguments are not yet settled. However, some of the criticisms of other theories may also apply to psychological essentialism – how might it help us understand complex concepts, for example?

SUMMARY OF SECTION 2

- The classical view, that concepts are definitions of categories, is undermined by arguments that many categories cannot be defined, and cannot readily explain typicality effects, borderline cases, and intransitivity.
- The prototype view, that categorization is determined by similarity to the prototype, explains most typicality effects. However, it cannot readily explain the context-sensitivity of typicality, nor how prototypes might combine in complex concepts. There is a residual question as to whether the existence of typicality effects implies a prototype organization.
- The theory-based view helps explain the non-independence of attributes in concepts, and dissociations between categorization and similarity. It also avoids some of the criticisms aimed at similarity. However, it is not clear how theories might combine in complex concepts, and the notion of a theory is very under-specified.
- Psychological essentialism apparently explains findings that even young children believe inner, hidden properties are causally responsible for category membership. However, it is not clear whether the notion of essence is required to explain data such as these. Moreover, the idea that people categorize according to essential properties has received mixed empirical support, as has the notion that people might defer to expert categorizations.

3 WHERE NEXT?

In this chapter we have canvassed some of the principal approaches that have been taken in developing a theory of concepts. In some respects, it seems as if the study of concepts is the study of theories that do not work for one reason or another. The classical view falters because we cannot identify necessary and sufficient conditions of category membership for all but a very few concepts. Prototype theory has difficulties explaining context-sensitivity and complex concepts. Ultimately, both suffer for their use of the notion of similarity, which seems unable to explain categorization fully. Theory-based notions of concepts are imprecise and cannot obviously explain complex concepts. Lastly, psychological essentialism has received mixed empirical support, and much of the empirical evidence only indirectly relates to the notion of essences.

However, such a picture of the psychology of concepts is unnecessarily gloomy. Indeed, it turns out that we have probably learned more about the phenomena of categorization even as various theories have been found wanting. And, of course, adherents of those theories continue to introduce modifications in order to explain recalcitrant data. Nonetheless, our discussion of the different theoretical approaches raises (at least) two questions. What sense can we make of so many different theoretical treatments when none is without problems? And where might researchers next turn their attention if there is, as yet, no common theoretical framework? As I shall try to suggest, one way of answering these questions is to consider to what extent categorization is a unitary phenomenon.

3.1 Is all categorization the same?

Perhaps the different theoretical treatments of concepts reflect the fact that categorization is not one single process. Maybe people categorize items in different ways in different circumstances. Indeed, discursive psychologists, whose approach we earlier bracketed off, might argue that categorization depends *essentially* on context, and that there is nothing common to all the cases that we call categorization. Were context to have such an unbridled influence we might expect categorization to appear unsystematic. Yet much of the evidence presented in this chapter points to the opposite – we have examined a wide range of empirical data that are highly robust.

One way of reconciling the idea that people categorize things differently on different occasions with the idea that categorization is nonetheless systematic is to suggest that there are (a determinate number of) different kinds of categorization. Moreover, it is conceivable that these could be usefully framed by the different theories of concepts. For example, perhaps the classical view gives a useful account of categorization in cases where we need to provide or appeal to definitions. In law, for instance, often we need to reach an agreement or adopt a convention as to whether something belongs to a category (e.g. whether a 16-year-old is a child or an adult). Similarly, prototype theory may usefully explain categorization in circumstances where we need to categorize something rapidly, or perhaps under uncertainty, maybe when we are in a position to take into account only an object's superficial properties. Likewise, theory-based views may describe categorization when we are seeking a more reflective and considered judgement, perhaps when we are using categorization in order to explain something. And essentialism may usefully explain how we categorize when we wish to be consistent with expertise and a scientific knowledge of the world.

Machery (2005) makes a similar point. In his examination of complex concepts, he argues that during categorization people employ a range of information, including elements of commonsense theories, exemplar-level information, and prototypes. If this is right, then it would be simplistic to expect one of the theories we have examined to provide a comprehensive account of concepts and categorization.

Speculative though this possibility is, Smith and Sloman (1994) have provided some empirical evidence that there may be some truth to it. They sought to replicate some of the dissociations between similarity and categorization judgements obtained by Rips (1989) and described in Section 2.4.2. Rips found that people judged an object intermediate in size between a quarter and a pizza to be more similar to a quarter, but more likely to be a pizza. Smith and Sloman obtained the same dissociation only when participants were required to think aloud whilst making their decisions and so articulate reasons for their judgements (that is, they provided a concurrent verbal protocol – see Chapter 10). Smith and Sloman interpreted this finding as pointing to two modes of categorization: a similarity-based mode of categorization and a rule-based mode. The implication is that people will either focus on similarity or on underlying rules or structure depending on how the categorization task is presented. When in similarity-based mode, categorization seems to conform to similarity-based accounts, such as prototype theories. When in rule-based mode, categorization seems to be more theory- or explanation-based. Though this does not show that there are as many different ways of categorizing as there are theories of concepts, it does suggest that categorization may not be a single process. It is a possibility, therefore, that some of the different accounts of concepts may be implicitly concerned with different kinds of purpose in categorization, and ultimately with different kinds of categorization.

In a similar vein we can rethink the phenomena that are taken as evidence of the nature of concepts. Earlier we noted that concepts and words bear a complex relationship to one another, but much of the evidence we have so far reviewed has tended to equate the use of category words with categorization. However, while our use of category labels is certainly influenced by our beliefs about categorization, it is also influenced by language more generally. Indeed, we can label something with a category word yet not believe that it belongs to the category – describing a statue of a lion as a 'lion', for example, does not indicate that we think the statue really is a lion. Malt *et al.* (1999) showed that the same is true for how we label containers, such as 'box', 'bottle', and 'jar'. They found that whether an item was called a 'bottle' depended not so much on how similar it was to a prototypical bottle, but whether there was something similar that was also called a 'bottle'. In this way, for example, a shampoo container might get called a shampoo 'bottle' despite bearing little similarity to a prototypical bottle. So, whether we apply a category label (e.g. bottle) to an object depends in part on how that label has been used historically and only in part on whether we think that the object really belongs to the labelled category (i.e. on whether the object really is a bottle).

3.2 Are all concepts the same?

Another possibility that we should consider is the extent to which different types of concept require a different theoretical treatment. Already you might have noticed how each theory seems to work most convincingly for a slightly different set of examples. In Activity 5.2 you tried to define a range of different categories: sparrow, gold, chair, introvert, red, and even number. Did you feel then that these categories were very different from one another? If so, we can perhaps make sense of this intuition.

Some categories like even number seem amenable to definition. For these well-defined categories, the classical view appears to give a good explanation of category membership, though it does not obviously explain how some even numbers are considered more typical than others. Perhaps this would require something like Armstrong *et al.*'s dual-process account, and involve its attendant difficulties (see Section 2.2.1). Nonetheless it may be that a modified classical view would provide a good explanation of these kinds of category.

In a similar vein, prototype theories seem to work well for fuzzy categories – like red in Activity 5.2 – categories that seem to have genuine borderline cases. For these similarity to a prototype might provide the best explanation of category membership, since there is no prospect of defining these categories and people in general do not seem to have relevant commonsense theories (e.g. a theory of the deeper causal principles by which red things come to appear red). Perhaps categories like chair are fuzzy in the same way.

Theory-based and essentialist approaches are likely to be most successful for categories for which people have commonsense theories. Perhaps unsurprisingly, these include many categories for which scientific theories have also been developed, e.g. sparrow and gold from Activity 5.2. These are categories where it is relevant to develop a deeper, explanatory knowledge of the causal principles underlying the category. As we have seen, however, researchers have extended the claims of essentialism to include other kinds of category, such as artefact and social categories, e.g. introvert in Activity 5.2.

While it remains only a possibility that different concepts may require different theoretical treatments, there is evidence nevertheless that different kinds of concepts may be realized in the brain in different ways (e.g. Warrington and Shallice, 1984; Strnad *et al.*, 2011; though see also Tyler *et al.*, 2003). This evidence may yet challenge the view that one theory could accommodate all kinds of concept. Even if people were to conclude that different categories require different theoretical treatments, it would still be important to find some way of relating the different theories so we could understand in what sense they were all theories of concepts.

3.3 Are all categorizers the same?

Consonant with above considerations, we might also consider whether all categorizers are the same. Medin *et al.* (1997) recruited participants from three occupational groups with correspondingly different experience and knowledge of trees: maintenance workers, landscapers, and taxonomists. They then asked them to sort the names of 48 different kinds of tree into whatever groups made sense. The taxonomists tended to reproduce a scientific way of sorting the trees; the maintenance workers produced a similar sorting, although they gave more emphasis to superficial characteristics (such as whether trees were broadleaved). They also tended to include a 'weed tree' group that was not present in the taxonomists' sorts, and which included trees that cause particular maintenance problems. The landscapers didn't reproduce a scientific taxonomy, but justified their sorts in terms of factors such as landscape utility, size, and aesthetic value. Lynch *et al.* (2000) also showed how the typicality ratings of the same kinds of tree expert differed from those of novices. Typicality for the expert group reflected similarity to ideals – i.e. trees judged to be best examples of the category were not of average or prototypical height, but of extreme height; in contrast, the ratings of the novices were largely influenced by familiarity.

Studies such as these suggest that different people do not necessarily categorize things in the same way. The goals that a person has as well as the extent of their knowledge may influence the way they categorize and, by extension, be reflected in their concepts.

It is also possible that categorization may differ in more fundamental ways for different kinds of person. For example, there may be deeper individual differences, possibly rooted in cognitive style (see, for example, the discussion of essentialism above). There is also evidence for disruption of semantic processing in schizophrenia (Chen *et al.*, 1994), and accumulating evidence that categorization is altered in autism (Coutinho *et al.*, 2010), Alzheimer's dementia (Doughty *et al.*, 2009), and schizotypy (Morgan *et al.*, 2009). Morgan *et al.*'s results are of special interest since, in this and similar studies, the sample of (relatively) high schizotypes are drawn from the normal, healthy population, typically students, and therefore differences among categorizers cannot easily be dismissed as manifestations of disordered cognition.

SUMMARY OF SECTION 3

It is possible that the failings of one or all approaches to concepts may be due to any combination of the following:

- Categorization may not be a single process, and different kinds of categorization may lend themselves to different theoretical treatments.
- Different types of category have different properties and so may require different theoretical treatments.
- Different groups of people may categorize things in different ways according to their goals and the nature of their knowledge, and so may fit the claims of different theories.

4 CONCLUSION

Overall, it seems that category knowledge is multilayered, encompassing knowledge of the causal properties relevant to a category, knowledge relevant to explaining category membership and the properties of instances, knowledge of function, and knowledge of superficial properties useful for identification and judgements about appearance. It also seems that we are capable of calling on different kinds of category knowledge on different occasions and for different purposes. While these observations are not inconsistent with a single theoretical treatment of concepts, they nonetheless raise the prospect that competing theories provide good explanations of somewhat different sets of phenomena, and so are not directly in contradiction. However the theoretical debates may or may not be resolved, I hope this chapter has convinced you of the importance of concepts to an understanding of cognition. Though categorization presents substantial challenges for researchers, these are challenges for all cognitive psychologists. Only once they have been met are we likely to be able to develop a good understanding of the mind. (None of which is likely to trouble Rosie.)

FIGURE 5.3 Rosie (untroubled).

Answers

Here are the identities of the objects shown in Figure 5.1: (a) is an olive stoner; (b) is an asparagus peeler; and (c) is a pickle picker, ideal for retrieving the very last pickled onion or gherkin from a jar.

FURTHER READING

Inevitably in a chapter of this length, I have omitted some important issues. Most notably, I have only briefly touched on the exemplar view of concepts, formal models of categorization, and the literature on category learning, and have not discussed the issue of basic level concepts. For these and other important aspects of concepts and categorization, I would strongly recommend Greg Murphy's excellent book. For a philosophically-inspired selection of psychological and philosophical works, see Laurence and Margolis.

Laurence, S. and Margolis, E. (eds.) (1999) *Concepts: Core Readings*, Cambridge, MA, MIT Press. Laurence and Margolis's reader contains an excellent choice of philosophical and psychological papers that together represent most key lines of thought and argument pertaining to concepts.

Murphy, G.L. (2002) *The Big Book of Concepts*, Cambridge, MA, MIT Press. Greg Murphy's book provides a highly readable and comprehensive view of the psychological literature on concepts and categorization.

REFERENCES

Allen, S. and Brooks, L. (1991) 'Specializing the operation of an explicit rule', *Journal of Experimental Psychology: General*, vol.120, pp.3–19.

American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, Washington, DC, American Psychiatric Association.

Armstrong, S.L., Gleitman, L., and Gleitman, H. (1983) 'What some concepts might not be', *Cognition*, vol.13, pp.263–308.

Barsalou, L.W. (1983) 'Ad hoc categories', *Memory & Cognition*, vol.11, pp.211–27.

Barsalou, L.W. (1985) 'Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.11, pp.629–54.

Bloom, P. (1996) 'Intention, history, and artifact concepts', *Cognition*, vol.60, pp.1–29.

Bloom, P. (2007) 'More than words: a reply to Malt and Sloman', *Cognition*, vol.105, pp.649–55.

Braisby, N.R. (2001) 'Deference in categorization: evidence for essentialism?' in Moore, J.D. and Stenning, K. (eds.) *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, Mahwah, NJ, Lawrence Erlbaum.

Braisby, N.R. and Franks, B. (1997) 'What does word use tell us about conceptual content?' *Psychology of Language and Communication*, vol.1, no.2, pp.5–16.

Braisby, N. and Hanlon, S. (2010) 'Categorization, deference and cognitive style', in Ohlsson, S. and Catrambone, R. (eds.) *Proceedings of the 32nd Annual Conference of the Cognitive Science Society,* Austin, TX, Cognitive Science Society.

Braisby, N. and Hodges, I. (2009) 'Categorization of sexual orientation: a test of essentialism', in Taatgen, N.A. and van Rijn, H. (eds.) *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, Austin, TX, Cognitive Science Society.

Braisby, N.R., Franks, B., and Hampton, J.A. (1996) 'Essentialism, word use and concepts', *Cognition*, vol.59, pp.247–74.

Brooks, L.R., Norman, G., and Allen, S. (1991) 'Role of specific similarity in a medical diagnosis task', *Journal of Experimental Psychology: General*, vol.120, pp.278–87.

Bruner, J.S., Goodnow, J.J., and Austin, G.A. (1956) *A Study of Thinking*, New York, John Wiley.

Chater, N. and Heyes, C.M. (1994) 'Animal concepts: content and discontent', *Mind & Language*, vol.9, pp.209–46.

Chen, E.Y.H., Wilkins, A.J., and McKenna, P.J. (1994) 'Semantic memory is both impaired and anomalous in schizophrenia', *Psychological Medicine*, vol.24, pp.193–202.

Coutinho, M.V., Volker, M.A., and Mercado, E. (2010) 'Atypical categorization in children with high functioning autism spectrum disorder', *Psychonomics Bulletin and Review*, vol.17, pp.862–8.

Coxon, A.P.M. (1999) *Sorting Data: Collection and Analysis (Quantitative Applications in the Social Sciences)*, Thousand Oaks, Sage.

Doughty, O.J., Lawrence, V.A., Al-Mousawi, A., Ashaye, K., and Done, D.J. (2009) 'Overinclusive thought and loosening of associations are not unique to schizophrenia and are produced in Alzheimer's dementia', *Cognitive Neuropsychiatry*, vol.14, no.3, pp.149–64.

Eco, U. (1999) *Kant and the Platypus: Essays on Language and Cognition*, London, Secker & Warburg.

Edwards, D. and Potter, J. (1992) *Discursive Psychology*, London, Sage.

Fodor, J.A. (1998) *Concepts: Where Cognitive Science Went Wrong*, Oxford, Clarendon Press.

Gellatly, A.R.H. (1997) 'Why the young child has neither a theory of mind nor a theory of anything else', *Human Development*, vol.40, pp.1–19.

Gelman, S. and Wellman, H. (1991) 'Insides and essences: early understandings of the non-obvious', *Cognition*, vol.38, pp.213–44.

Goodman, N. (1972) 'Seven strictures on similarity', in Goodman, N. (ed.) *Problems and Projects*, Indianapolis, IN, Bobbs-Merrill.

Gopnik, A. (1996) 'The scientist as child', *Philosophy of Science*, vol.63, 485–514.

Hahn, U., Prat-Sala, M., Pothos, E.M., and Brumby, D.P. (2010) 'Exemplar similarity and rule application', *Cognition*, vol.114, pp.1–18.

Hampton, J.A. (1982) 'A demonstration of intransitivity in natural categories', *Cognition*, vol.12, pp.151–64.

Hampton, J.A. (1998) 'Similarity-based categorization and fuzziness of natural categories', *Cognition*, vol.65, pp.137–65.

Hampton, J.A., Estes, Z., and Simmons, S. (2007) 'Metamorphosis: essence, appearance and behavior in the categorization of natural kinds', *Memory & Cognition*, vol.35, pp.1785–800.

Haslam, N. and Levy, S. (2006) 'Essentialist beliefs about homosexuality: structure and implications for prejudice', *Personality and Social Psychology Bulletin*, vol.32, pp.471–85.

Haslam, N. and Whelan, J. (2008) 'Human natures: psychological essentialism in thinking about differences between people', *Social and Personality Psychology Compass*, vol.2/3, pp.1297–312.

Haslam, N., Rothschild, L., and Ernst, D. (2000) 'Essentialist beliefs about social categories', *British Journal of Social Psychology*, vol.39, pp.113–27.

Haslam, N., Rothschild, L., and Ernst, D. (2002) 'Are essentialist beliefs associated with prejudice?' *British Journal of Social Psychology*, vol.41, pp.87–100.

Hull, C.L. (1920) 'Quantitative aspects of the evolution of concepts', *Psychological Monographs*, vol.28.

Jylkkä, J., Railo, H., and Haukioja, J. (2009) 'Psychological essentialism and semantic externalism: evidence for externalism in lay speakers' language use', *Philosophical Psychology*, vol.22, no.1, pp.37–60.

Keil, F. (1989) *Concepts, Kinds and Cognitive Development*, Cambridge, MA, MIT Press.

Kripke, S.A. (1972) 'Naming and necessity', in Davidson, D. and Harman, G. (eds.) *Semantics of Natural Languages*, Dordrecht, Reidel.

Kroska, A. and Goldstone, R.L. (1996) 'Dissociations in the similarity and categorization of emotions', *Cognition and Emotion*, vol.10, no.1, pp.27–45.

Lynch, E.B., Coley, J.D., and Medin, D.L. (2000) 'Tall is typical: central tendency, ideal dimensions and graded category structure among tree experts and novices', *Memory and Cognition*, vol.28, no.1, pp.41–50.

Machery, E. (2005) 'Concepts are not a natural kind', *Philosophy of Science*, vol.72, pp.444–67.

Macintyre, A. (1998) *ME: Chronic Fatigue Syndrome, A Practical Guide*, London, Thorsons.

Malt, B.C. (1990) 'Features and beliefs in the mental representation of categories', *Journal of Memory and Language*, vol.29, pp.289–315.

Malt, B.C. (1994) 'Water is not H_2O', *Cognitive Psychology*, vol.27, pp.41–70.

Malt, B.C. and Sloman, S.A. (2007) 'Category essence or essentially pragmatic? Creator's intention in naming and what's really what', *Cognition*, vol.105, pp.615–48.

Malt, B.C., Sloman, S.A., Gennari, S., Shi, M., and Wang, Y. (1999) 'Knowing versus naming: similarity and the linguistic categorization of artefacts', *Journal of Memory and Language*, vol.40, pp.230–62.

McCloskey, M. and Glucksberg, S. (1978) 'Natural categories: well-defined or fuzzy sets?' *Memory and Cognition*, vol.6, pp.462–72.

Medin, D.L. (1989) 'Concepts and conceptual structure', *American Psychologist*, vol.44, no.12, pp.1469–81.

Medin, D.L. and Ortony, A. (1989) 'Psychological essentialism', in Vosniadou, S. and Ortony, A. (eds.) *Similarity and Analogical Reasoning*, Cambridge, CUP.

Medin, D.L. and Schaffer, M.M. (1978) 'Context theory of classification learning', *Psychological Review*, vol.85, no.3, pp.207–38.

Medin, D.L. and Shoben, E.J. (1988) 'Context and structure in conceptual combination', *Cognitive Psychology*, vol.20, pp.158–90.

Medin, D.L., Lynch, E.B., Coley, J.D., and Atran, S. (1997) 'Categorization and reasoning among tree experts: do all roads lead to Rome?' *Cognitive Psychology*, vol.32, pp.49–96.

Minda, J.P. and Smith, J.D. (2011) 'Prototype models of categorization: basic formulation, predictions, and limitations', in Pothos, E. and Wills, A. (eds.) *Formal Approaches in Categorization*, Cambridge, UK, Cambridge University Press.

Morgan, C.J.A., Bedford, N.J., O'Regan, A., and Rossell, S.L. (2009) 'Is semantic processing impaired in individuals with high schizotypy?' *Journal of Nervous & Mental Disease*, vol.197, no.4, pp.232–8.

Murphy, G.L. (2000) 'Explanatory concepts', in Keil, F.C. and Wilson, R.A. (eds.) *Explanation and Cognition*, Cambridge, MA, MIT Press.

Murphy, G.L. and Medin, D.L. (1985) 'The role of theories in conceptual coherence', *Psychological Review*, vol.92, pp.289–316.

Newen, A. and Bartels, A. (2007) 'Animal minds and the possession of concepts', *Philosophical Psychology*, vol.20, no.3, pp.283–308.

Nosofsky, R.M. (2011) ‘The generalised context model: an exemplar model of classification’, In Pothos, E. and Wills, A. (eds.) *Formal Approaches in Categorization,* Cambridge, UK, Cambridge University Press.

Pinker, S. (1997) *How the Mind Works*, London, Penguin Books.

Potter, J. and Wetherell, M. (1987) *Discourse and Social Psychology*, London, Sage.

Putnam, H. (1975) ‘The meaning of “meaning”’, in Putnam, H. (ed.) *Mind, Language and Reality, Volume 2: Philosophical Papers*, Cambridge, CUP.

Rips, L.J. (1989) ‘Similarity, typicality and categorization’, in Vosniadou, S. and Ortony, A. (eds.) *Similarity and Analogical Reasoning*, Cambridge, CUP.

Rips, L.J. and Collins, A. (1993) ‘Categories and resemblance’, *Journal of Experimental Psychology: General*, vol.122, pp.468–86.

Rips, L.J., Shoben, E.J., and Smith, E.E. (1973) ‘Semantic distance and the verification of semantic relations’, *Journal of Verbal Learning and Verbal Behaviour*, vol.12, pp.1–20.

Roberson, D., Davidoff, J., and Braisby, N. (1999) ‘Similarity and categorization: neuropsychological evidence for a dissociation in explicit categorization tasks’, *Cognition*, vol.71, pp.1–42.

Rosch, E.H. (1973) ‘On the internal structure of perceptual and semantic categories’, in Moore, T.E. (ed.) *Cognitive Development and the Acquisition of Language*, New York, Academic Press.

Rosch, E.H. (1975) ‘Cognitive representations of semantic categories’, *Journal of Experimental Psychology: General*, vol.104, pp.192–233.

Rosch, E.H. and Mervis, C.B. (1975) ‘Family resemblances: studies in the internal structure of categories’, *Cognitive Psychology*, vol.7, pp.573–605.

Ross, B.H. and Murphy, G.L. (1999) ‘Food for thought: cross-classification and category organization in a complex real-world domain’, *Cognitive Psychology*, vol.38, pp.495–553.

Roth, E.M. and Shoben, E.J. (1983) ‘The effect of context on the structure of categories’, *Cognitive Psychology*, vol.15, pp.346–78.

Sappington, B.F. and Goldman, L. (1994) ‘Discrimination learning and concept formation in the Arabian horse’, *Journal of Animal Science*, vol.72, no.12, pp.3080–7.

Smith, E.E. and Sloman, S.A. (1994) ‘Similarity- versus rule-based categorization’, *Memory & Cognition*, vol.22, no.4, pp.377–86.

Smith, E.E., Osherson, D.N., Rips, L.J., and Keane, M. (1988) ‘Combining prototypes: a selective modification model’, *Cognitive Science*, vol.12, pp.485–527.

Strevens, M. (2000) ‘The essentialist aspect of naïve theories’, *Cognition*, vol.74, pp.149–75.

Strnad, L., Anzellotti, S., and Caramazza, A. (2011) ‘Formal models of categorization: insights from cognitive neuroscience’, in Pothos, E. and Wills, A. (eds.) *Formal Approaches in Categorization,* Cambridge, UK, Cambridge University Press.

Sutcliffe, J.P. (1993) ‘In defence of the “classical view”: a realist account of class and concept’, in van Mechelen, J., Hampton, J.A., Michalski, R., and Theuns, P. (eds.), *Categories and Concepts: Theoretical Views and Inductive Data Analysis*, London, Academic Press.

Tyler, L.K., Bright, P., Dick, E., Tavares, P., Pilgrim, L., Fletcher, P., Greer, M., and Moss, H. (2003) ‘Do semantic categories activate distinct cortical regions? Evidence for a distributed neural semantic system’, *Cognitive Neuropsychology*, vol.20, no.3/4/5/6, pp.541–59.

Warrington, E.K. and Shallice, T. (1984) ‘Category specific semantic impairments’, *Brain*, vol.107, pp.829–54.

Witkin, H.A., Dyk, R.B., Faterson, H.F., Goodenough, D.R., and Karp, S.A. (1962) *Psychological Differentiation: Studies of Development*, New York, John Wiley and Sons, Inc.

Wittgenstein, L. (1953) *Philosophical Investigations*, Oxford, Basil Blackwell, translated by G.E.M. Anscombe.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

CHAPTER 6

LANGUAGE PROCESSING

Gareth Gaskell and Helen Brown

1 INTRODUCTION

What are the qualities of human beings that differentiate us from other species? You can probably think of many characteristics, but pretty high on most people's lists would be the ability to produce and understand language. Linguistic abilities underpin all manner of social interactions – from simple acts such as buying a bus ticket or greeting a friend, right up to constructing and refining political and legal systems. Like many aspects of cognition, the ability to use language develops apparently effortlessly in the early years of life, and can be applied rapidly and automatically.

This chapter looks under the surface of the language system, in order to understand the unconscious operations that take place during language processing. Our focus is on the basic mechanisms required for language understanding. For example, understanding a simple spoken sentence involves a whole string of abilities: the perceptual system must be able to identify speech sounds, locate word boundaries in sentences, recognize words, access their meanings, and then integrate the word meanings into a coherent whole, respecting the grammatical role each word plays. Each of these abilities has been extensively researched, with numerous models proposed to explain each aspect of language processing; this chapter provides an overview of our current understanding of how language is processed. As you will see, there often remains considerable disagreement about some quite fundamental properties of the language system. Nonetheless, there has also been substantial progress in terms of identifying some of the features required of the language system for it to work the way it does.

The building blocks of language identified in this chapter are discussed in a wider context in Chapter 7, which examines, for example, questions such as how speakers interact in conversation. Chapter 7 also covers language production, whereas the current chapter concentrates on language *perception*. The structure of this chapter roughly follows the time course of processing in language perception. Section 2 builds on some of the ideas about recognition introduced in Chapter 4, but looks specifically at the processes that result in the identification of spoken and written words. Models of these processes generally assume that word recognition involves access to a mental lexicon – something that was briefly introduced in Chapter 5 in the context of lexical concepts – that stores relevant information relating to the words we know (e.g. what they mean). Section 3 deals with the contents of the mental lexicon and how this information might be organized. Finally, Section 4 looks at the process of sentence comprehension beyond the mental lexicon. It deals with how listeners use their knowledge of the grammar of a language to construct the meaning of a sentence. In each section some of the influential models of language processing are discussed, along with key experimental studies that help us to evaluate and refine these models.

2 WORD RECOGNITION

Adult speakers of English tend to know somewhere between 50,000 and 100,000 words. Most common words are easy to describe and use, suggesting that their meanings are clearly accessible. Less common words are perhaps represented more vaguely, with some words difficult to define out of context, but nonetheless generating a feeling of familiarity. For example, you might be reasonably confident that *tarantella* is a word and have good knowledge of how it should be pronounced, but you might still be unable to give a good definition of what it is (a fast whirling dance, once believed to be a cure for a tarantula bite!).

So, quite a lot of information is stored in the mental lexicon about word meanings and pronunciations. The goal of word recognition is to access this information as quickly as possible. We shall look at how this process occurs in two different sensory modalities: auditory and visual. This may at first seem repetitious, but there are some important differences between the two modalities that, at this level of the language system, lead to quite different models of recognition processes. Before you read through the sections on word recognition, you may wish to remind yourself of the broader issues involved in recognition, as described in Chapter 4.

2.1 Spoken word recognition

Speech is the primary medium of language. Widespread literacy has emerged only in some cultures, and only in the last century or two, meaning that reading is, in evolutionary terms, a new ability. Speech in contrast is something that almost all humans acquire, and has been around long enough for some aspects of spoken language to be thought of as innate. Speech is also primary in the sense that we learn to understand and produce speech before we learn to read and write. For these reasons, we will firstly look at how spoken word recognition operates, and then go on to examine the visual modality.

2.1.1 Segmenting the speech stream

Our conscious experience of spoken words is in some ways similar to our experience of text on a page: words are perceived as coherent and discrete events, so we generally don't experience any difficulty in finding the dividing line between two words. However, the truth of the matter is that the speech waveform has no simple equivalent of the white space between printed words. Instead, as Activity 6.1 shows, silent gaps are unreliable as indicators of spoken word boundaries. Yet somehow the language system must be able to divide the speech stream up so that the words contained in it can be recognized and understood. How then does this speech segmentation process operate?

Models of segmentation can generally be divided into two types: (1) *pre-lexical* models and (2) *lexical* models. Pre-lexical models rely on characteristics of the speech stream that might mark a likely word boundary, whereas in lexical models segmentation is guided by knowledge of *how words sound*. The first model is pretty straightforward: the only issue at stake

✱ ACTIVITY 6.1

The waveform in Figure 6.1 depicts a typical sound wave that might enter the ear when you hear someone speak. Try to work out from the sound wave how many words have been spoken, and pencil in a mark where you think each word boundary lies.

COMMENT

When you have noted down your estimates, compare them with the actual boundaries marked at the end of the chapter. How did you decide on likely word boundaries, and was this method a useful one? Most people assume that silent gaps between words are likely boundary markers, but they can be misleading. Some word boundaries do not involve silence because the surrounding **phonemes** are coarticulated, meaning that they blend together. A phoneme is the speech equivalent of a letter (they are normally annotated with surrounding slash marks), so, for example, /k/ and /ə/ are the first two phonemes in *confess*. **Coarticulation** refers to the fact that you have to prepare for upcoming phonemes well before they are produced, and these preparations lead to changes in the phonemes currently being pronounced. For example, the /d/ phoneme in ‘do’ and ‘dah’ sounds slightly different because of the following vowel. In addition, some silent gaps do not mark word boundaries: they are just points where the airways are closed in the course of uttering a word. For example, when you say the word ‘spoken’ your lips close briefly in order to produce the sudden release of air in the phoneme /p/. This results in a short period of silence between the /s/ and the /p/.

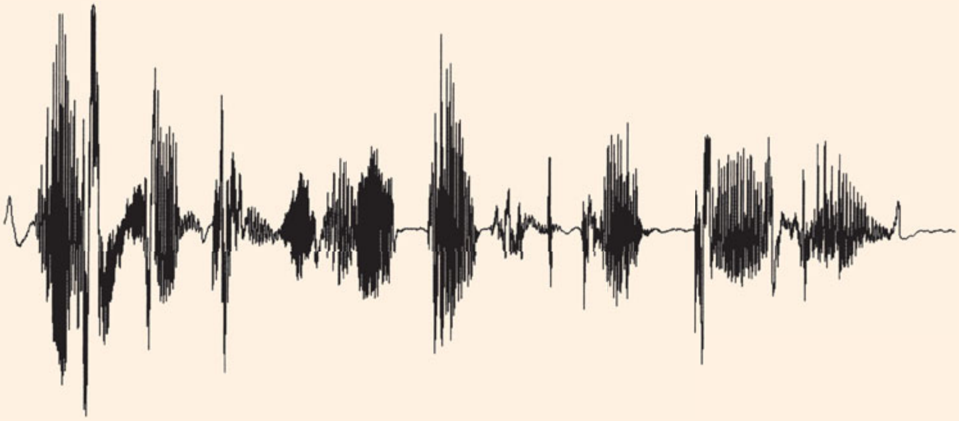


FIGURE 6.1 A speech waveform.

is what type of characteristic or cue can be extracted from the speech waveform as a useful indicator of a word boundary. We have already seen that silent gaps are not sufficient, but nonetheless silence can be useful, particularly if it lasts quite a long time.

Another important pre-lexical cue comes from the *rhythm* of speech. All languages have some unit of temporal regularity, and this provides the basic rhythm when an utterance is produced. In English, this unit is known as a **metrical foot**, and consists of a strong (stressed) syllable, followed optionally by one or more weak (unstressed) syllables (as you can see from Figure 6.2). Strong syllables are reasonably clear landmarks in the speech stream, and most words that have a meaning (such as *bacon* or *throw*) rather than a grammatical role (e.g. *it*, *of*) begin with a strong syllable (Cutler and Carter, 1987). So, a segmentation strategy that predicts a word boundary before each strong syllable would seem like a valuable one for English speakers.

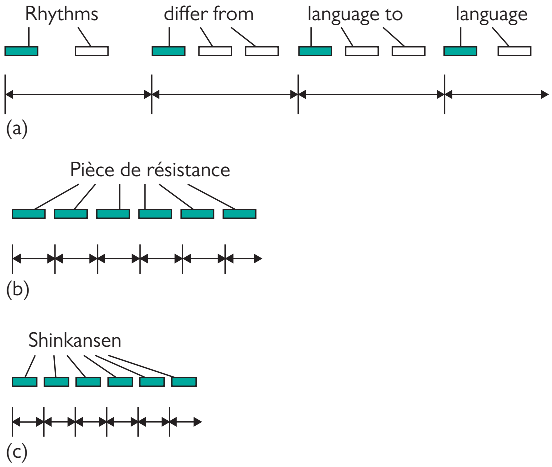


FIGURE 6.2 Examples of language rhythms. In English (a) the basic unit of rhythm is the strong syllable (the filled boxes). These stressed syllables are roughly equally spaced out in time when you produce a sentence, no matter how many weak syllables (unfilled boxes) there are between the strong syllables. Each group of strong and weak syllables is known as a 'foot', so when you say the sentence in (a), it may feel like you are speaking more quickly towards the end of the sentence because you need to fit in more weak syllables to maintain the gaps between the strong syllables. In French (b), the syllable is the unit of rhythm and so all syllables are roughly equally spaced in time. The rhythmic unit in Japanese (c) can be even smaller than a syllable. For example, shinkansen ('bullet train') contains six units of rhythm (including three single consonants), but only three syllables.

Cutler and Norris (1988) provided evidence supporting this idea: they played pairs of nonsense syllables to listeners and asked them to monitor for any familiar word embedded in the speech (this is known as the wordspotting task – think trainspotting but duller). For example, in the sequence 'mintayve', which consists of two strong syllables, there is the embedded word 'mint'. Cutler and Norris argued that for a sequence like this listeners should identify the two strong syllable onsets (the /m/ and the /t/), and search for any words they know beginning at those points. This segmentation would obscure recognition of the word 'mint' because it spans a hypothesized word boundary (i.e. they would tend to hear two units: 'min' and 'tayve'). On the other hand, a sequence like 'mintesh' (where the second weak syllable contains a reduced 'uh' vowel) would tend to be segmented as a single unit, and so spotting the word 'mint' should be relatively easy. Their prediction turned out to be correct, suggesting that listeners make use of the rhythm of English in order to identify likely word boundaries. In languages where different rhythmic units dominate, such as French (syllables) or Japanese (sub-syllabic units), similar sensitivities have been demonstrated (see Cutler and Otake, 2002), suggesting that early in life people 'tune into' their native language and optimize their segmentation strategy accordingly.

These (and other) pre-lexical cues are clearly valuable for identifying likely word boundaries in a sentence of utterances. However, none of the models that rely on pre-lexical cues can claim complete accuracy in boundary identification. This means that there will be cases where a boundary is incorrectly predicted, and other cases where a real boundary is missed. For example, a word like *confess* begins with a weak syllable, and so its onset would be missed by a pre-lexical segmentation strategy based on strong syllables. It seems that there must be some other mechanism available for cases like this. This is where lexical models can offer more insight: lexical segmentation models rely on our knowledge of particular words' phonological representation (what they sound like) to guide segmentation. The simplest version of this kind of strategy would involve recognizing each word in an utterance sequentially, and so predicting a new word at the boundary of the existing word (e.g. Marslen-Wilson and Welsh, 1978). For example, think about how the sentence 'Confess tomorrow or die!' might be segmented. If you can recognize the first word quickly (before it finishes), then you can use the knowledge that this word ends in /s/ to predict a word boundary as soon as the /s/ is encountered. You can then start again on word recognition with the speech following the /s/ (*tomorrow*). The problem here though is that most words are much shorter than *confess* and *tomorrow*, and cannot be recognized within the time-span of their acoustic waveforms, meaning that a lot of backtracking would be required to locate word boundaries using this method (think about trying to segment the sentence 'Own up now or die' using the same strategy).

We shall return to this issue in Section 2.1.3, when we evaluate the TRACE model of spoken-word recognition (McClelland and Elman, 1986), which provides

BOX 6.1 RESEARCH STUDY Learning to segment speech

The cross-linguistic differences between segmentation mechanisms highlight the fact that the ability to segment speech is one that must be learned during the course of language development. French and English babies aren't innately specified with different segmentation mechanisms; instead these develop as a consequence of exposure to language. Saffran *et al.* (1996) provided an impressive demonstration of how statistical information can aid the development of both segmentation and vocabulary acquisition.

They devised an artificial language made up of three-syllable words such as *dapiku* or *golatu*, and then used a computer to synthesize a long continuous stream of speech containing these 'words'. Their intention was to produce a sequence in their artificial language that contained absolutely no acoustic or rhythmic cues to the location of word boundaries. If people only make use of acoustic and rhythmic segmentation cues then the speech they hear should appear as unsegmented nonsense. However, if they can make use of statistical information about *co-occurrence* of syllables, then they may start to pick out the words of the language. In other words, they might start to notice that the syllables *da*, *pi*, and *ku* quite often occur in sequence.

Using what is known as a 'head-turning' procedure, 8-month-old infants were tested on their perception of this kind of speech. The infants were presented with words from the artificial language on one loudspeaker and jumbled syllables (e.g. *pikugo*) on another. The idea was that if the infants found the words from the language familiar, they might spend more time listening to the novel sequences (and so turn their heads towards the associated loudspeaker). Using this technique, Saffran *et al.* (1996) found that the infants *did* begin to pick out the words from the stream of syllables after just two minutes of the speech. This ability to learn the statistical properties of patterns is quite universal – it operates for adults and children as well as babies, and works just as well for non-speech stimuli such as tones or shapes (Saffran *et al.*, 1999). Therefore, speech segmentation may make use of a wide-ranging **implicit learning** ability, which may even be shared by other primates, such as tamarin monkeys (Hauser *et al.*, 2001).

Interestingly, Cunillera *et al.* (2010) demonstrated that when adults learned some of the 'words' at the beginning of the experiment, before exposure to a speech stream containing both the newly learned words and six additional unlearned words, they were better at segmenting the unlearned words compared to when the speech stream did not contain any previously learned words. This finding suggests that adults make use of multiple cues, including lexical knowledge and statistical information, when segmenting speech.

a more powerful lexical-segmentation mechanism. Although there is plenty of evidence supporting pre-lexical mechanisms it remains likely that lexical competition operates alongside them to provide a more robust system for dividing up the speech stream.

2.1.2 Parallel activation

A spoken word typically lasts about half a second. In many ways it might simplify matters if the recognition process began only once the whole of a word had been heard. However, for the language system this would be valuable time wasted. Instead, speech is continually evaluated and re-evaluated against numerous potential candidates in order to determine the identity of each word: this is known as **parallel activation**. A great advantage of this method of assessment is that it can lead to determination of a word's identity well before the end of the word is heard.

The mechanism sketched above is most clearly exemplified by the cohort model of Marslen-Wilson and colleagues (Marslen-Wilson and Welsh, 1978; Marslen- Wilson, 1987). This model assumes that as the beginning of a word is encountered, the **word-initial cohort** (a set of words that match the speech so far) is activated. For example, if the beginning of the word were 'cuh' (as in *confess*), then the word-initial cohort would include words like *canoe*, *cocoon*, *karate*, and so on, because these words all match the speech so far. Then, as more of the word is heard, the recognition process simply becomes one of whittling down the set of potential candidates. For example, 'conf . . .' would rule out all the words above, but not *confess*, *confetti*, or *confide*. At some point in this process (the **uniqueness point**) the candidate set should be reduced to a single word. According to the cohort model, the recognition process is then complete. As mentioned above, the recognition point in this kind of model can be well before the end of the word, meaning that valuable time is saved in interpreting the speaker's message. However, not all words become unique before the end of the word; for example, '*cat*' can stand on its own as a word, but can also form the first syllable in a number of longer words, such as *caterpillar, catalogue*, and *catch*. Thus, even when the entire word has been heard there will still be a number of active potential candidates. In this case, the word *cat* will only be recognized after the listener has heard part of the following word and determined that this does not constitute a legal continuation of the word *cat*.

Nevertheless, even this conception of the process doesn't reflect the full fluency of word recognition. So far, we haven't discussed the *goal* of the recognition process – accessing our stored knowledge about a word. One might assume that this occurs at the recognition point of a word. However, it seems that access to meaning can occur substantially earlier. Marslen-Wilson (1987) demonstrated this using **cross-modal priming**. This technique – which is used to examine the extent to which the meaning of a spoken word has been retrieved – involves hearing a spoken **prime** word, followed swiftly by a visual **target** word. Participants were given the task of deciding whether the target was a word or not as quickly as possible. Semantic similarity between a prime–target pair such as 'confess' and *sin* leads to faster responses to the target (compared with an unrelated control pair, such as 'tennis' and *sin*). This implies that, on reaching the end of the word, the meaning of 'confess' has been activated. The question that Marslen-Wilson addressed was whether the meaning would be activated at an earlier point, before the uniqueness point had been reached. He found that when something like 'confe . . .' was used as a prime, responses to the target word *sin* were still facilitated. The same spoken fragment would also facilitate responses to the target *wedding*, which was semantically related to an alternative cohort member, *confetti*. This suggests that the meanings of both *confess* and *confetti* are briefly accessed while the word *confess* is being heard.

You might want to reflect on what this result means in terms of how we recognize spoken words. It suggests that when we hear a word, we don't just activate the meaning of that word; we also activate, very briefly, the meanings of other words that begin with the same phonemes. Meanings of likely candidates are activated before the perceptual system can identify the word being heard, which ensures that the relevant meaning has been retrieved by the time the word is identified.

Parallel activation of multiple meanings is an important property of the language recognition system. The alternative – a serial search, which would be a bit like looking through a dictionary for a word meaning – is unlikely to be as efficient (particularly for words near the end of the list). However, it is worth questioning the extent of parallel activation. For example, could it be the case that there is no limit to the number of meanings that can be activated briefly? And can these multiple meanings be accessed without any interference between them? Gaskell and Marslen-Wilson (2002) argued that meaning activation is limited, again on the basis of cross-modal priming data. They showed that if many meanings are activated at the same time, the resultant priming effect is relatively weak compared with the amount of priming found when just one or two meanings are compatible with the speech input. It appears to be the case that activating more than one meaning can only occur partially, so the gradual reduction of the cohort set of matching words is accompanied by a gradual isolation and amplification of the relevant meaning. Nonetheless, an overriding characteristic of the speech perception system is to initially access *too much* information rather than too little. This maximizes the chances of having accessed the correct meaning as soon as enough information has been perceived to identify the particular word.

2.1.3 Lexical competition

Marslen-Wilson's cohort model was important because it incorporated parallel evaluation of multiple lexical candidates, and emphasized the swiftness and efficiency of the recognition process. Later models used a slightly different characterization and relied on the activation and competition metaphor (introduced in Chapter 4 in the discussion of the IAC model of face recognition). In these models, each word in the lexicon is associated with an **activation level** during word recognition, which reflects the strength of evidence in favour of that particular word. The cohort model in its original form (Marslen-Wilson and Welsh, 1978) can be thought of as a dichotomous activation model: words are either members of the cohort (equivalent to an activation level of 1) or they aren't (activation level 0). The advantage of more general models of lexical competition such as TRACE (McClelland and Elman, 1986) is that they can use continuously varying activation levels to reflect the strengths of hypotheses more generally. This is useful in cases where a number of words are consistent with the speech input so far, but the information in the speech stream matches some words better than others. If activation levels are on a continuous scale, then this inequality can be reflected in the activations assigned to word candidates.

The TRACE model is a connectionist model that assumes three levels of representation: the **phonetic feature** level (phonetic features describe the component parts of how a phoneme is articulated, such as lip closure or tongue position), the phoneme level, and the word level (containing a node for each word the listener knows). The idea of the model is that the speech stream is represented as changing patterns of activation at the phonetic feature level. These nodes feed into a phoneme recognition level, where a phonemic representation of speech is constructed. A word node has connections from all the phonemes within that word. For example, the *confess* node would have connections feeding into it from the /k/, /ə/, /n/, /f/, /E/, and /s/ phoneme nodes. If the phoneme nodes for that word became activated, activation would then spread to the *confess* word node, resulting in strong activation for that word. The net result is that word-node activations reflect the degree to which each word matches the incoming speech.

A second mechanism provides a way of selecting between active words. Nodes at the word level in TRACE are connected by inhibitory links. When any word node becomes activated, it starts to inhibit all other word nodes (i.e. by decreasing their activation) with the strength of inhibition depending on the degree to which that node is activated. This competitive element tends to amplify differences in word activations, so that it becomes clear which words are actually in the speech stream and which are just *similar to* the words in the speech stream. So if the spoken word was *confess*, then the node for *confetti* would become strongly activated as well, because all phonemes in the input apart from the final one fit the representation of *confetti*. However, the *confess* node would be activated to a slightly greater extent because all phonemes in the input are consistent. Both these word nodes would be strongly inhibited by the other, but the greater **bottom-up support** (i.e. greater consistency with the incoming signal) for *confess* would ensure that the *confess* node would eventually win the competition, remaining activated when the *confetti* node had been strongly inhibited.

This 'winner-takes-all' activation and competition approach is common to many models both within language (we shall see another example in Section 2.2) and across cognition (e.g. face recognition). These commonalities across different areas of cognition are valuable, as they provide a way of extracting more general principles of cognitive processing from specific examples. In the case of speech, lexical competition provides a simple mechanism for deciding which words best match the speech input. As we saw in Section 2.1.1, it also provides a subsidiary means of segmenting the speech stream into words. This is because it is not just words that have the same onset, such as *confess* and *confetti*, that compete, but also words that simply overlap to some extent, such as *confess* and *fester* (see Figure 6.3). These words have a syllable in common (i.e. the second syllable of confess and the first syllable of fester). If this syllable is perceived, then both of these words will become activated, but through lexical competition only one will remain active. The segmentation problem can then be

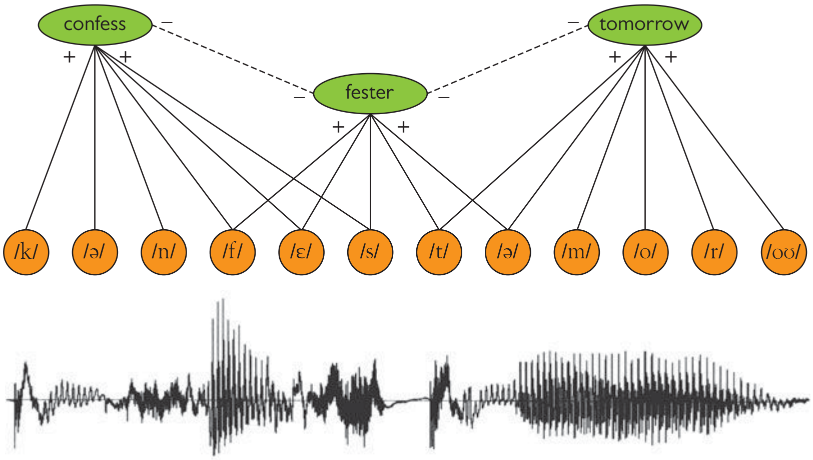


FIGURE 6.3 Illustration of lexical competition in the TRACE model. The speech stream activates a set of phonetic feature nodes (not shown), which then activate the corresponding phoneme nodes. Word nodes at the lexical level are linked up to the relevant phoneme nodes with positive links (solid lines). In this case, the speech is actually 'confess tomorrow'. This sequence actually fits three words completely: confess, fester, and tomorrow. These word nodes have inhibitory connections between them that vary in strength depending on their degree of overlap. So there is no inhibitory link between the confess and tomorrow nodes, but the fester node has inhibitory links to both confess and tomorrow (broken lines). The combined inhibition of the fester node from the other nodes has the effect of suppressing its activation, leaving only confess and tomorrow as active lexical candidates. The competitive links between words in TRACE allow word recognition to be carried out, and also provide a mechanism for word boundary identification (there must be a boundary between the /s/ of confess and the /t/ of tomorrow).

viewed as having been solved *implicitly* in the activation of the word nodes. For example, if *confess* wins the competition then there must be a word boundary at the end of the syllable '*fess*', but if fester wins the competition then the boundary must be at the start of '*fess*'.

This general version of lexical competition is supported by a wordspotting experiment by McQueen *et al.* (1994). They looked at the time taken to spot a word like *mess* in two different types of embedding sequence. In a sequence like 'duhmess' the first two syllables match a longer word: *domestic*. If lexical competition operates for all overlapping words (see Figure 6.3), then the inhibitory link from the *domestic* node should make it difficult to spot *mess*. McQueen *et al.* found that detection rates were indeed lower and slower in this case compared to a case like 'nuhmess', in which a longer competitor does not exist. Interestingly, Bowers *et al.* (2009) found that the degree to which short embedded words such as *mess* are activated can vary depending on the position of the embedded sequence within a longer word. Using a picture-word interference task, in which participants name a picture while listening to a semantically related distractor word, embedded words only appeared to interfere with picture naming if the embedded word was aligned with a syllable boundary (e.g. *bat* in *acrobat*), not when the embedded word was misaligned with a syllable boundary (e.g. *ear* in *pioneer*). As such, it appears that the processes involved in lexical competition interact with speech segmentation mechanisms during spoken word recognition, and that these two processes may be carried out in parallel.

2.2 Visual word recognition

In this section we focus on the special qualities of word recognition in the visual domain, looking at how the recognition process operates, how visual and auditory processes are linked, and how eye movements are linked to the recognition system. Compared to speech, text might be thought of as an unproblematic medium. After all, it is relatively easy to spot where words begin and end, and text isn't transient in the way that speech is – if you misperceive a word on the page, you can simply go back to that word and try again. However, the availability of textual information also raises specific issues that must be addressed by models of visual word recognition. For example, because textual information is freely available over an extended period of time, we need to understand how the recognition system determines where the eyes should fixate, and for how long.

2.2.1 Models of visual word recognition

We have already seen how TRACE models *spoken* word recognition in terms of activation and competition in a multi-level connectionist network. TRACE was in fact a variant of an earlier model of visual word recognition proposed by McClelland and Rumelhart (1981). The visual model is often known as the IAC (interactive activation and competition) model, and shares many properties with the IAC model of face recognition you met in Chapter 4. The model contains three levels of nodes, representing activation of (1) visual features, (2) letters, and (3) words. Like TRACE, it is the inhibitory units within a level that provide a competitive activation system. Visual input is represented by activation at the featural level, and facilitatory and inhibitory links between levels of representation allow activation to build up at the higher levels. In this way, visual word recognition can be modelled as an interactive competition process.

An important property of many of these competition networks is that as well as allowing activation to flow up through the system (i.e. from features through letters to words), during the course of recognition they also allow activation to flow in the other direction (from words downwards). This is another example of the concept of top-down processing that was introduced in Chapter 3. For example, if the word node for *slim* became activated, the activation would feed back through facilitatory links to the constituent letter nodes (i.e. 's', 'l', 'i', and 'm'). At first glance, these feedback links appear redundant, because the letter nodes are going to be activated in any case by bottom-up sensory information. But their value becomes apparent in cases where the bottom-up information is degraded in some way. For example, suppose that the first letter of *slim* was obscured slightly, so that the 's' letter node was only weakly activated by the visual input. In this case, the *slim* word node would still be activated by the three unambiguous letters, and would in turn increase the activation of the 's' letter node. The result would be correct recognition of the obscured letter, despite the weakness of the sensory evidence.

This kind of top-down influence can be useful in explaining **lexical effects** on lower-level processing. A classic finding in word recognition (known as the 'word superiority effect', or WSE) is that letter detection is easier when the letter forms part of a word (e.g. the letter 'i' is easier to detect in *slim* than in *spim*). This can be attributed to the influence of the word node for *slim* providing a secondary source of activation for recognition of 'i', whereas there is no secondary source for a non-word like *spim*. So the top-down feedback connections in the IAC model provide a neat explanation of why we often find lexical influences on recognition of sub-lexical units like letters.

However, Grainger and Jacobs (1994) demonstrated that a variant of the IAC model could also explain the WSE without any top-down feedback. They proposed that responses to the letter-detection task were based on two different levels of representation: a letter-detection response could be based on activation of letter nodes *or* word nodes. The idea here was that one of the pieces of information about a word stored in the mental lexicon is a description of the written form of the word. So if a word node reaches a critical level then the spelling of that word should be activated, triggering a response. The upshot was that the incorporation of a second basis for responses using lexical information allowed the WSE to be accommodated in a model that only used bottom-up flow of activation.

The experimental finding of WSE remains a robust and important phenomenon, but the research of Grainger and Jacobs shows that there is more than one way of explaining the effect. Whether or not top-down processing is needed is one of the most contentious questions in the area of word recognition (both auditory and visual) and other areas of perception, and it remains a hotly debated topic amongst cognitive psychologists (e.g. Norris *et al.*, 2000, and associated commentaries).

2.2.2 Mappings between spelling and sound

So far we have treated the question of how words are recognized separately for spoken and written words. This section looks at how these two modalities interact, and what this tells us about the language system. There is an obvious need for interaction in order to spell a word that you have just heard, or to read aloud a written sentence. But there are more subtle reasons for suspecting that there are links between the **orthography** of a word (its spelling) and its **phonology** (its sound). Some of the data we shall now look at suggest that visual word recognition relies strongly on spoken word representations and processes.

As described in Activity 6.2, reading aloud is often portrayed in terms of two separate mechanisms: assembled and addressed phonology. The separate mechanisms are explicitly represented in dual-route models of reading such as the DRC model of Coltheart *et al.* (2001). DRC (see Figure 6.4) is a complex and powerful model, and builds on more than 100 years of theorizing about multiple routes in reading processes. For current purposes, the critical feature of the model is that it contains a 'rule-based' route to pronunciation via a grapheme–phoneme rule system (assembled phonology; see right-hand side of Figure 6.4), plus a 'lexical' route that requires retrieval of a stored pronunciation (addressed phonology; see left-hand side of Figure 6.4). Looking at the speed with which written words can be named often assesses the degree to which these routes are involved in reading. A typical finding is that regular words (e.g. *bell*) are named faster than irregular words (e.g. *pint*), but that this advantage is only present for **low-frequency** words (i.e. words that occur relatively rarely in the language). This can be explained by dual-route models in terms of a race between the two routes. Regular words can be named via either the lexical or the rule-based route to pronunciation, whereas irregular words can only make use of the lexical route. On the whole, naming speeds are faster when two routes are available (naming a regular word can be based on the output of whichever route delivers the pronunciation first) than when only one route is available (irregular

✱ ACTIVITY 6.2

Think about what processes might operate when you read aloud the following words: *bell*, *stick*, *pint*, *yacht*, *colonel*. How does a reader convert the orthographic (or written) form to a phonological one in order to pronounce the words, and where is the phonological information stored? Would the same processes operate when you read the following non-words: *dobe*, *leck*, *brane*, *noyz*?

COMMENT

Researchers often refer to two different ways of reading words aloud (what we could call 'retrieving their phonology'). The division is much like the division between phonics and whole-word methods of teaching children to read. *Assembled phonology* (like phonics) means generating a pronunciation based on a set of mappings between letters and sounds for your language. For example, the 'b' in *bell* corresponds to the /b/ phoneme, and there are similar conversion rules for 'e' and 'll'. This works well for words such as *bell* and *stick*, because they follow these conversion rules (i.e. they are regular items), and also for non-words like *dobe* and *leck*. But what about *pint*, *yacht*, and *colonel*? A simple sounding out of these irregular words would lead to the wrong pronunciation (e.g. *pint* might be pronounced to rhyme with *hint*), suggesting that an alternative mechanism is available. This is often known as *addressed phonology*, and (like whole-word methods of teaching reading) it relies on some kind of stored pronunciation of the whole word in the mental lexicon. *Brane* and *noyz* are unusual because their pronunciation coincides with the pronunciations of real words (i.e. *brain* and *noise*). These **pseudohomophones** (non-words that can be pronounced to sound like words) are generally only found in rock lyrics and some rather fiendish language experiments (see below).

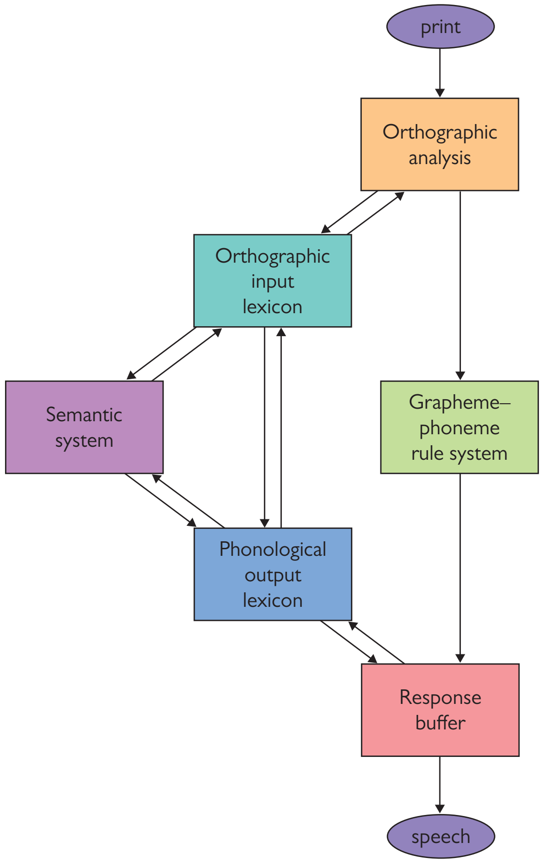

print
Orthographic analysis
Orthographic input lexicon
Semantic system
Grapheme–phoneme rule system
Phonological output lexicon
Response buffer
speech

FIGURE 6.4 DRC model. *Source: Coltheart et al., 2001, Figure 6*

words). For low-frequency words, the advantage of two routes over one for regular words results in them being named more quickly. For high-frequency words, it is assumed that the lexical route operates very quickly regardless of regularity, and so the influence of the additional rule-based route is minimal.

Regularity is not the only variable that determines the speed with which a word can be named. Glushko (1979) showed that the properties of neighbouring words – words that have similar spellings, *not* words that are in neighbouring locations – are also critical. For example, people are quick to name a word like *wade*, because it is a regular word, but also because all neighbouring words with the same final letters have a consistent pronunciation (e.g. *made*, *jade*, *spade*). On the other hand, although wave is also a regular word, its neighbours are inconsistent in terms of pronunciation (e.g. *have* and *slave* don't rhyme). This inconsistency results in slower naming. Simple dual-route models, which apply the same rules to regular items irrespective of their neighbours, could not easily explain differences between words containing consistent spelling-to-sound mappings (e.g. -*ade*), and words containing inconsistent spelling-to-sound mappings (e.g. -*ave*). Partly in response to this issue, Seidenberg and McClelland (1989) proposed that a single connectionist network could provide a basis for modelling naming of both regular and irregular items, while accounting for effects of neighbouring items, as in Glushko's consistency effect.

There is an obvious need for phonology to be accessed in reading aloud, because speaking requires a phonological representation. But does phonology also have a role to play when a reader simply has to identify and understand written words? Van Orden (1987) showed that phonological representations are involved in silent reading even when they are detrimental to performance. Van Orden asked participants to decide whether visually presented words were members of particular categories, such as whether a *rose* is a flower. Critically, participants found it difficult to reject homophones such as *rows* when making these decisions. In these cases participants would frequently make an incorrect response, suggesting that they were activating the pronunciation of the homophone and this was creating confusion. A similar effect was found when the critical items were pseudohomophones (e.g. *roze*).

Other studies have shown that spoken word representations are heavily involved in visual word recognition in many different languages. This may seem rather bizarre – surely word recognition based on visual features would be simpler and quicker? But we need to remember that speech perception is to some extent an innate ability, and we learn to understand spoken language very early in life. So when we begin to read, we already have a perfectly tuned recognition system for speech. It therefore makes sense for the visual recognition system to 'latch onto' the spoken system in order to ease the learning process.

What is perhaps more surprising is that visual word representations appear to be involved in spoken word recognition. One of the first demonstrations of this comes from a study by Seidenberg and Tanenhaus (1979) who asked participants simply to decide whether two words rhymed or not. They found that participants were able to answer this question faster when the two words were spelled in a similar way (e.g. *tie/pie*) compared to when the two words were spelled differently (e.g. *tie/guy*). More recent studies have provided evidence suggesting that the activation of orthographic information during spoken word recognition occurs automatically. Chéreau *et al.* (2007), for example, found that spoken target words were processed faster when they overlapped with the spoken prime in both phonology and orthography (e.g. *dream/gleam*) compared to when the prime and target overlapped only in phonology (e.g. *scheme/gleam*), and that this was the case even when the gap between hearing the prime and target was reduced to 20 ms. This finding demonstrates that even when participants have little time to process a spoken word, orthographic information is still activated and affects processing of the spoken item. It seems, therefore, that all forms of information are used when faced with the task of recognizing a word, be it information about how the word sounds or about how the word is spelled. More importantly, this appears to be the case regardless of whether the word is presented in a spoken or a written form.

2.2.3 Eye movements in reading

Speech perception is a relatively passive process, in that the listener doesn't need to perform any overt action in order to listen to a conversation. Reading a book or newspaper, however, is more active, because the reader controls the speed of uptake of information and must direct their eyes in order to take in new information. Eye movements turn out to be enormously useful in revealing how the language system operates.

Eye movements may feel quite smooth and continuous introspectively, but they really consist of **saccades** (jerky movements) followed by **fixations** (more-or-less stationary periods) during which visual information is processed. Eye-tracking techniques can monitor the movements of the eyes during reading

BOX 6.2 METHODS Eye tracking

Eye-tracking techniques generally rely on the fact that various parts of the eye such as the lens and the cornea reflect light. If a light source (usually infrared) is directed at the eye from a given angle, the angle of the reflection can be used to determine the orientation of the eye, and consequently the direction of gaze. Precise measurements can be made if the eye-tracking system combines measurements from more than one surface within the eye.

In studies of reading, the position of the head is often fixed using a chinrest and headrest and the participant is presented with text on a computer screen. Given that the head position is fixed and the distance from the screen is known, the reader's *gaze* location relative to the text can be calculated from the gaze-angle measurements. This results in a set of timed fixations to the text, as illustrated in Figure 6.5.

It is clear how eye-tracking studies would be beneficial for understanding how we read. However, a less obvious use of eye tracking is in the study of *spoken* language. Here, the participant is presented with a spoken sentence in the context of some visual scene, and the eye movements of the listener are monitored. For example, if a participant is sitting in front of a table with some candy and a candle on it, and is asked to pick up the candy, it is revealing to find out at what point people look at the candle, and correlate this with the amount of speech information they have received at that point (Tanenhaus *et al.*, 1995). In this kind of situation (see Section 4.4 for another example of this method), the participant needs to be able to move their head freely. To allow for this, a slightly different type of tracker is used, consisting of an eye tracker mounted on the head plus a second system for determining head position.

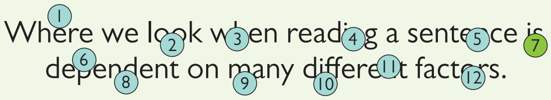

Where we look when reading a sentence is
dependent on many different factors.
1
2
3
4
5
6
7
8
9
10
11
12

FIGURE 6.5 Example of typical eye movements during reading.

and relate them to the location of the reader's gaze (see Box 6.2). Figure 6.5 illustrates the fixations involved in the processing of a typical sentence. Each numbered circle corresponds to the gaze location for a single fixation. Fixations typically last about 200 ms, but their durations are strongly dependent on the linguistic processing involved. For example, fixation duration is strongly dependent on the frequency of a word's usage in the language (Rayner and Duffy, 1986). This, along with many other effects, suggests that fixations are a measure of some kind of processing difficulty, and so they can reveal influential variables in reading.

As Figure 6.5 illustrates, our eyes don't simply move from one word to the next as reading proceeds. Some words are skipped altogether, whereas others require multiple fixations. In a significant proportion of cases, readers perform regressive saccades (i.e. they move backwards through the text), as marked by the grey circle in Figure 6.5. Short **function words** (grammatical words like *we* and *on*) are much more likely to be skipped than **content words** (words that convey meaning, like *sentence* and *look*), and regressions can often tell us about cases where a word has been misinterpreted due to some ambiguity (Starr and Rayner, 2001).

Eye-movement data are also valuable in terms of understanding where we fixate *within* a word. O'Regan and Jacobs (1992) showed that words are identified most quickly if they are fixated at a point in the word known as the **optimal viewing position OVP**. The OVP is generally near the middle of a word but can be slightly left of centre in the case of longer words. The fact that fixations work best if they are near the middle of the word makes sense, given that visual acuity is best in the **foveal** (central) region of the retina (try fixating on the edge of this page and reading the text!). The slight but consistent bias in favour of left of centre is more intriguing. Shillcock *et al.* (2000) argued that this bias reflects a balancing of the informativeness of the parts of the word to the left and right of the fixation point. The OVP should be left of centre for longer words because there is greater redundancy towards the end of most of these words. For the word *cognition*, for example, it would be easier to guess what the word is from the first five letters (*cogni*) than the last five (*ition*). Shillcock *et al.* also found that for some shorter words such as *it*, the theoretical OVP was outside the word, either to the left or the right, perhaps explaining why shorter words are often not fixated when reading.

How much textual information can be utilized when a reader makes a fixation? The foveal region of the retina has the greatest acuity, but it spans a limited angle. It is possible that information can also be gained from the **parafoveal** region, which is wider but has reduced acuity. Rayner and colleagues have carried out a number of ingenious experiments aimed at assessing the perceptual span of readers. In one experiment they applied a moving window to text shown on a computer screen so that only a certain number of letters to the left and right of the current fixation point could be read (see Figure 6.6). Whenever the participant shifted their gaze, the window shifted accordingly. They found that if the window is small, reading

```
xxxxx wexxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx   (3)
      *
xxxxxxxxxxxxxxxxxxxxxxding xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx   (5)
                        *
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxtence depxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx   (9)
                                    *
xxxxxxxxxxxok when readingxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx   (15)
                  *
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxentence depends on many dxxxxxxxxxxxxxxxx   (25)
                                          *
```

FIGURE 6.6 Example stimuli in a moving window experiment (e.g. McConkie and Rayner, 1975). The numbers in parentheses are the window sizes in characters. In a typical experiment the participant sits in front of a computer screen, with an eye-tracking system monitoring their gaze. Wherever the participant directs their gaze along a line of xs, the computer displays around that point small 'windows' of unobscured text. All text outside this window is obscured by xs (in some experiments the gaps between words were preserved). The asterisk below each example marks the fixation point, and would not be seen by a participant (participants can fixate where they like). In this case the sentence is 'Where we look when reading a sentence depends on many different factors'.

is a slow and painful process, but for larger windows participants are barely aware of the text beyond the window that is masked.

The idea behind this technique is that the experimenter can gradually increase the window size until a point is reached at which reading speed and comprehension are normal. At this point one can be confident that the text beyond the window range is not used for normal reading. The results suggest that the perceptual span for English readers is quite limited: about 15 characters to the right of fixation and 3 characters to the left. The asymmetry is due to the left-to-right nature of reading in English – there is more useful information to be gained in text following the fixation position than the text preceding the fixation (which has normally already been read). In a right-to-left language like Hebrew the asymmetry is swapped, showing that the perceptual span in reading is shaped by the requirements of the written language.

SUMMARY OF SECTION 2

- In speech, finding out where words begin and end is a non-trivial problem. Models of word segmentation rely on either features of the speech stream or knowledge about how words sound.
- Word recognition relies on parallel assessment of multiple options, and competition between word candidates.
- Models of word recognition differ in the extent to which top-down processing is required.
- Visual word recognition relies to a considerable extent on speech codes, whilst spoken word recognition relies to some extent on written codes.
- Studying eye movements during reading reveals what aspects of visual word recognition cause difficulties.

3 THE MENTAL LEXICON

In Section 2, word recognition was largely viewed as an identification procedure. That is to say, we assumed that the mental lexicon stores representations of what words sound and look like, and that when we hear or see a word there is a recognition process that compares the input with stored representations. However, identification is just the first step towards understanding a word – what we *really* need to know is what a word means. In this section we shall move beyond identification and look more closely at how word meanings are accessed during the recognition process. We shall look at the *semantic content* (how word meanings are stored) and the *semantic organization* (how word meanings are related) of the mental lexicon. Before we do this, we shall also take a brief look at quite a difficult area of language processing known as morphology.

3.1 Morphology

Morphology deals with the size of units in the mental lexicon. It's often taken for granted that the basic unit of the mental lexicon is the word, but in fact many words can be broken down into morphemes (the smallest meaningful unit within a word) when they are perceived. This implies that the morpheme is the true basic unit. This is particularly the case for languages such as Turkish in which words tend to be rather long, cumbersome units, and a lexicon based on morphemes would be more economical (fewer entries) and more flexible.

Most people are aware that words can be divided up into meaningful units, and that these units follow some grammatical rules. For example, we know that plurals in English are generally derived by adding an ‘s’ to the singular form, as in *cats* or *dogs*. In its spoken form, the rule is slightly more complex: speakers add on /s/, /z/, or /iz/, depending on the final phoneme of the singular form (think about how you would say the plural forms *cats*, *dogs*, and *pieces*). This kind of relatively minor modification of a word (for example, marking pluralization or tense) is known as an inflectional change, and is covered by a branch of morphology known as **inflectional morphology**. More major modifications are possible as well, in which the grammatical class of a word may change. For example, the suffix *-ness* can change an adjective to a noun (as in *happiness* or *weakness*). Similarly, the suffix *-ly* can change an adjective into an adverb. These modifications form part of **derivational morphology**.

The examples of morphological change given above are particularly straightforward. All involve regular changes in which the meaning of the word is predictable from the meanings of the morphemes. But things are not always so simple. For example, according to the regular pluralization rule, the plural form of *mouse* should be *mouses* not *mice*. *Mice* is an example of an irregular plural form, and similar irregularities exist in many types of morphological change. Similarly, the meanings of the morphemes making up a word may not always determine the meaning of the whole word. It’s easy to spot the relationship in meaning between *govern* and *government* but not between *depart* and *department*, yet both pairs have (at least supposedly) the same morphological relationship.

The descriptions given here are linguistic ones, but do they have any relevance to cognitive psychology? In other words, do the regularities that exist between families of words have any effect on the organization of the mental lexicon? It is quite possible that our recognition system is set up to recognize words, regardless of their sub-structure. This *full-listing approach* would mean that recognizing a word made up of many morphemes, such as *disenchantment*, is essentially the same process as recognizing a single-morpheme word, such as *cat*. The opposite extreme – often known as the affix-stripping or *decompositional approach* (Taft and Forster, 1975) – is that words are chopped up into morphemes as they are perceived, and the morpheme is the basic unit of representation in the lexicon.

One way of looking at whether the lexicon is organized in terms of morphemes is to test whether we can add morphemes onto unknown words. A classic demonstration of this kind of generalization is Berko’s ‘wug test’ (1958), in which children were encouraged to generate the plural form of novel words. For example, a child might see a drawing of a toy and be told that it was a *wug*. The child would then see two of these toys and be prompted to say what they were. Children found it easy to generate the correct inflected form (*wugs*), suggesting that they had learned some kind of pluralization rule, and that the ‘s’ can operate as an independent morpheme. The pluralization suffix is a particularly common one, but other morphemes such as *-ment*, as in *government*, or *en-*, as in *enact*, are more rare. This factor may affect the way different morphemes are represented in the lexicon. It may be that common morphemes such as the plural morpheme are stored as separate units, but less common units have no separate representation.

Marslen-Wilson *et al.* (1994) used the priming methodology to examine whether morphemic units exist in the mental lexicon. Their reasoning was that if words are broken down into morphemes then we should be able to get strong priming effects between words containing the same morpheme. They found that priming of this type depended on some shared meaning between the two words. So hearing *cruelty* resulted in faster processing of *cruel* because they have similar meanings, but hearing *casualty* did not prime responses to *casual*, presumably because there was no clear link between the meanings of the two words. However, more recent studies have shown that when the prime is presented for only a very short duration (typically around 40–50 ms), and participants are not aware of having seen the prime word, processing of the target word is speeded up even when there is no semantic link between the prime and target (Rastle *et al.*, 2004).

Taken together, these results suggest that at an early stage of processing words are automatically decomposed into their constituent morphemes irrespective of the meanings of the words, but that meaning may play a role during later processing stages. Thus, it seems that extreme positions such as full-listing or full decomposition are untenable. Perhaps it may instead make more sense to have a pragmatic view of morphological processing, in which morphological decomposition only occurs if there is some clear benefit to be had. Different morphemes within a language may be treated in different ways, and there may also be differences between languages in terms of the extent to which the mental lexicon relies on morphemes.

3.2 Accessing word meanings

Chapter 5 introduced you to the notion of lexical concepts – a class of concepts specific to words. In this section we shall relate the ideas underlying concepts and categories to the operation of the semantic system. We shall examine the kinds of information that become available once a word has been recognized, and also look at the problem of how to select the appropriate meaning in cases where a word is ambiguous.

3.2.1 Semantic representations

Once a word has been recognized the relevant information about that word must be accessed, so that the word, and ultimately the sentence, can be understood. Most models of language perception start to get slightly hazy at this point, because while word forms are quite concrete and easy to define, their meanings are rather less tangible and may vary quite strongly from person to person.

Two theories of how word meanings might be represented have gained popularity since the 1970s, both of which have links to the kinds of ideas discussed in Section 2.2 with respect to interactive activation models. Spreading activation models (e.g. Collins and Loftus, 1975) assume that words can be represented by units or nodes, as in the TRACE and IAC models of word recognition. The difference here is that links between nodes in spreading activation networks represent semantic relationships between words. Collins and Loftus's original model in 1975 used different kinds of links for different kinds of semantic relationship. For example, the network could encode the fact that a canary is a bird by linking the nodes for *canary* and *bird* with an 'is a' link, or that a canary has wings using a 'has' connection. Other models didn't use labelled links but simply connected together words that were similar in meaning. The application to word recognition would be that once a word has been recognized (for example by activating the correct node in the IAC model), activation would spread to the semantic network and then along links to related words, thus generating a set of known facts about that word and activating a set of semantically related words.

The alternative featural theory of semantic representation assumes that word meanings are represented as a set of semantic features or properties (a bit like some of the theories of concepts explored in Chapter 5). The idea here is that the mental lexicon contains a large set of features, and that each word representation consists of a subset of these features. For example, the features relevant for the word canary might include ('has wings', 'can fly', 'is a bird', and so on). The feature model has also been incorporated into connectionist models of recognition, allowing the linkage of recognition models and semantic representations. In this case, the activation of a written or spoken representation would lead to a pattern of activation on a set of semantic nodes, with each node representing a semantic feature (e.g. Masson, 1995).

These two approaches are highly underspecified, and could potentially accommodate many different patterns of data. Despite this, you might find it useful when you read through the experimental findings listed below to think about how the findings might be accommodated by featural and spreading activation theories. Most studies of semantic representations of words have addressed what kinds of information can be accessed and when. Clearly, all kinds of information about a word could be stored in the mental lexicon, but the information required to understand a sentence must be readily available in a fraction of a second, and this time constraint may have some consequences for what types of information are stored.

The most popular tool for investigating the types of semantic information stored in the mental lexicon is semantic priming. For example, an experiment might use pairs of semantically related words, such as *bread* and *butter*, with participants asked to perform some kind of speeded task such as lexical decision (is it a word or not?) or naming (say the word aloud) to the second item of the pair. In this case, the assumption is that if responses are facilitated (i.e. quicker) when there is a semantic relationship between the words, then that semantic relationship must be represented in the mental lexicon; in a spreading activation model there might be a link between the words, whilst in a featural model the words might share a number of semantic features.

So what kinds of relationship between words can support semantic priming? Perhaps the most robust effect involves pairs of **associated** words (words that seem to go together naturally). Association strength is often measured by asking people to say or write down the word that first comes into their heads when they read a target word. So if you were asked to provide an associate for the word *cheddar* you would probably say *cheese*. According to the University of South Florida norms (Nelson *et al.*, 1998), that's what more than 90 per cent of respondents say (curiously, a further 3 per cent of their respondents said *Swiss*!). In any case, the fact that presenting one word results in facilitated processing of an associated word suggests that associative links between words are represented in the lexicon in some way.

The problem with this conclusion is that the types of relationship found for associated word pairs are quite variable, ranging from near synonyms (words that have very similar meanings, such as *portion* and *part*), to antonyms (words that have opposite meanings, such as *gain* and *lose*), to words that just crop up in the same context (e.g. *law* and *break*). For this reason, researchers have often tried to look for semantic priming in cases where words have only weak associations but still retain some specific semantic link (e.g. *horse* and *sheep*). The data here are less clear, which suggests that non-associative links might be weaker in some way, or rely on a different mechanism compared with associative links. Nonetheless, Lucas (2000) reviewed a large set of semantic priming experiments and reached the conclusion that non-associative semantic priming effects were robust, with perhaps the strongest evidence for links in the lexicon between members of the same category (e.g. *horse–pig*) and instrument–action pairs such as *broom* and *sweep*.

Kellenbach *et al.* (2000) looked at whether words might be linked in terms of the visual or perceptual properties of the objects they represented. For example, *button* and *coin* both refer to flat, round objects. This kind of priming had been observed weakly in some studies but not others. However, Kellenbach *et al.* (2000) used two measures of priming: the first was the standard reaction time test, and the second was based on brain activity using the ERP technique (see Box 6.3 in Section 3.2.2). They found no effect in the reaction time test, but nonetheless a robust effect on the brain response to the target word, suggesting that even in this case, where the semantic link was too subtle to be detected by conventional techniques, a priming relationship still existed. So it seems that the semantic information that becomes available when a word is perceived is far from minimal. Instead, many different aspects of meaning are accessed. Current research says little about how these different aspects of meaning are organized and accessed, but even at this stage it seems that associative, pure semantic, and perceptual knowledge might be accessed in different ways.

3.2.2 Semantic ambiguity

In many cases, the operation of activating a word's meaning in the mental lexicon is made more difficult because the word is ambiguous in some way. For example, what does the word *bank* mean to you? You may immediately think of a high-street bank, but then later realize that *bank* could mean the side of a river as well. This is because *bank* is a homonym: a word that has multiple unrelated meanings. There are also more subtle possibilities: the first meaning of bank is most commonly applied to the place you keep your money. But a blood bank, while clearly related, is a somewhat different concept, as is the bank at a casino. So *bank* is a polysemous word, as well as a homonym, because it has multiple related senses. Further ambiguity is caused by the fact that bank could be a verb (transitive or intransitive) or a noun, but we shall leave this **syntactic ambiguity** to the next section. Homonyms are

BOX 6.3 METHODS Event-related potential (ERP) studies of semantic processing

The ERP method relies on the fact that brain activity creates an electromagnetic field that can be measured by a set of electrodes placed on the scalp. Typically, the recording of activity is synchronized with the presentation of a stimulus, and many recordings using different stimuli must be averaged to generate an interpretable waveform. The resultant ERP waveform often contains a set of characteristic peaks at different delays.

A negative peak occurring roughly 400 ms after the stimulus has been presented (known as the N400) has been identified with the integration of semantic information into sentential context. A typical finding is that the size of the N400 peak associated with a word in sentential context is inversely related to how easily that word fits into the context (Kutas and Hillyard, 1980). So the N400 peak associated with the word *spoon* might be small in the sentence 'James ate the cereal with a dessert spoon', but large in the sentence 'James caught the salmon using a fishing spoon'. This sensitivity to semantic congruency makes the ERP technique an excellent one for examining issues such as lexical ambiguity resolution.

Van Petten and Kutas (1987) compared ERP and standard priming methods of assessing the effects of sentential context on meaning activation for ambiguous words such as *bank*. They showed that even when standard priming techniques detected no influence of sentential context, the ERP waveforms for the ambiguous words were subtly different, suggesting that sentential context was influencing the processing of these words, and strengthening the case for an interactive account of lexical ambiguity resolution.

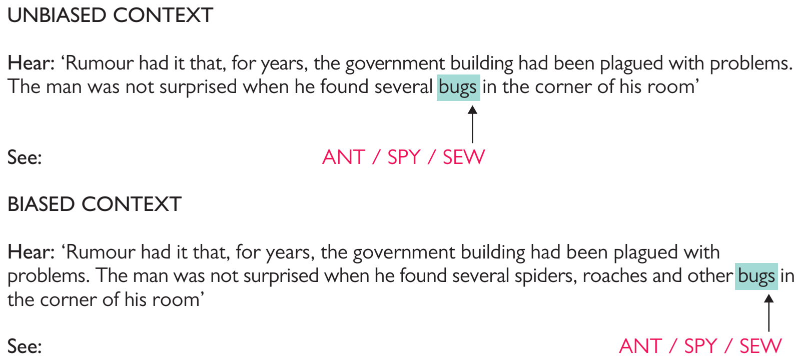
thankfully reasonably rare (roughly 7 per cent of common English words according to Rodd *et al.*, 2002), but the vast majority of words have multiple senses, which means that we really need to deal with ambiguity effectively if we are going to understand language.

Normally, the sentential context of an ambiguous word will provide some valuable clues to allow the relevant meaning of the word to be selected. So the question that researchers have focused on is *how* sentential context influences meaning selection in cases of ambiguity. Two opposing views have emerged since the 1980s (you may note similarities between the debate here and the debate on top-down and bottom-up processing discussed in Section 2.2). According to the *autonomous* view, all meanings of an ambiguous word are first accessed and then the contextually compatible meaning is selected from these alternatives. The *interactive* view has a stronger role for sentential context, in that it may in some cases rule out inappropriate meanings before they are fully accessed. So these two viewpoints differ in terms of whether there is a short period of time in which meanings of words are accessed regardless of sentential context.

Using cross-modal semantic priming, Swinney (1979) found evidence for autonomous activation of ambiguous word meanings. In his experiment, participants heard homonyms like 'bugs' embedded in sentential contexts, and were asked to make a lexical decision on a visual target related to one of the meanings of the prime or an unrelated control word (see Figure 6.7).

Swinney found that whether or not the sentence context was biased towards one meaning of the homonym, both related targets were primed. This implies that both meanings of the ambiguous word were accessed, despite the fact that in the biased condition only one meaning was compatible with the sentential context. When the experiment was repeated with the targets presented roughly one second later, only the contextually appropriate meaning appeared to be activated. So Swinney's data suggested that there is a short window of up to a second in which the meanings of ambiguous words are accessed without regard to sentential context, supporting the autonomous model.

Variants of Swinney's experiment have been run many times, and once again there is some inconsistency in the pattern of priming. In some cases it seems



UNBIASED CONTEXT

Hear: 'Rumour had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several bugs in the corner of his room'

See: ANT / SPY / SEW

BIASED CONTEXT

Hear: 'Rumour had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several spiders, roaches and other bugs in the corner of his room'

See: ANT / SPY / SEW

FIGURE 6.7 Example trial in Swinney's (1979) priming experiment. In the unbiased context, both meanings of bugs are plausible (relating to insects and relating to spying). The activation of each meaning is assessed using the reaction time to a related word (ant or spy), compared with a control unrelated word (sew). In the biased context, only the insect meaning is plausible by the time the homonym is heard.

that only one meaning is activated if the homonym has one particularly common meaning and the sentential context is strongly constraining towards that meaning (Tabossi and Zardon, 1993). Lucas (1999) has also shown that studies demonstrating exhaustive access of ambiguous word meanings often still show *more* priming for the contextually appropriate meaning than the inappropriate one. Therefore it seems that at least some interactive processing is likely in accessing word meanings, although sentential context may only rule out inappropriate meanings in specific circumstances.

SUMMARY OF SECTION 3

- The mental lexicon stores the meanings of words. Although the subject is contentious, it seems that some words are broken down into smaller units called morphemes.
- A wide variety of information about the meaning of a word becomes available when a word is recognized, including associative knowledge, pure semantic information, and perceptual features.
- For words with more than one meaning, the sentential context of the meaning can help select the relevant meaning. This process is to some extent interactive.

4 SENTENCE COMPREHENSION

So far, language perception has largely been described in terms of recognition processes. Up to the level of the lexicon, the job of the perceptual system is simply to allow recognition of familiar sequences (words or morphemes) and retrieve stored knowledge relating to these items. When we discussed morphology, there was a little more **productivity** involved. That is, people can recognize and make use of novel morphological variants of familiar morphemes. So, for example, even if the word *polysemous* was new to you when it was mentioned in the previous section, you would probably find it quite easy to define its morphological relative, *polysemy*. However, when we get to the level of the sentence, the character of language perception changes abruptly. Sentences are almost always new, in that the same permutation of words has often never been encountered before. If perception at this level were still simply a recognition process, then we would completely fail to understand all but the most simple or common sentences. The solution to this problem is to treat sentence-processing not as a pure recognition process but as a *constructive* process. When we read or hear a sentence, we take the individual components – the words – and combine them to produce something that may be quite novel to us, but hopefully bears some relationship to the message the speaker or writer intended. You might think of this process in terms of building up a mental model of the information being communicated (see Chapter 12 on reasoning). Accordingly, the listener or reader takes each word and deduces its grammatical or syntactic role in the current sentence. Termed **parsing**, this process is the focus of the final section in this chapter.

4.1 Syntax

Before embarking on a review of the models and data relevant to sentence processing, it is worth having a quick look at linguistic views of language structure.

✱ ACTIVITY 6.3

Please read the following passages and sentences and think about whether they seem grammatical to you. Give each one a rating from 1 to 10, where grammatical-sounding passages get high marks.

1. The most beautiful thing we can experience is the mysterious. It is the source of all true art and all science. He to whom this emotion is a stranger, who can no longer pause to wonder and stand rapt in awe, is as good as dead: his eyes are closed.
2. Her five-year mission: to explore strange, new worlds; to seek out new life and new civilizations; to boldly go where no man has gone before.
3. Please cup, gimme cup.
4. Colourless green ideas sleep furiously.
5. In become words sentence the rather have jumbled this.

COMMENT

People have quite reliable intuitions about the grammaticality of sentences, despite often being unable to define exactly what makes a sentence grammatical. You probably gave the first two passages fairly high ratings. Passage 1 is a quotation from Albert Einstein, and applies quite well to the study of syntax: mysterious but potentially very revealing! Passage 2 may be familiar as the opening line of the *Star Trek* series. You might be tempted to mark this down as being less grammatical because it contains a famous example of a split infinitive: 'to boldly go'. However, this kind of (most likely mistaken) grammatical rule is not what cognitive psychologists are typically interested in; we do not wish to dictate what the best or most eloquent way of speaking is, we simply wish to understand how people really speak. In these terms splitting the infinitive is a perfectly acceptable and grammatical form of language. Sentence 3 is not grammatical by most definitions, but if a two-year-old said it to you, you would understand what they meant quite easily. Sentence 4 is in some ways the opposite of Sentence 3, in that it seems grammatical, yet meaningless. It was made famous by Noam Chomsky as an example of how syntax and semantics can be dissociable. Finally, Sentence 5 is clearly ungrammatical and pretty hard to extract any meaning from. After a while you may be able to work out that the sentence is a scrambled version of 'The words in this sentence have become rather jumbled'. It demonstrates just how important it is for us to have some mutually agreed conventions for word order, and this is precisely what syntax is!

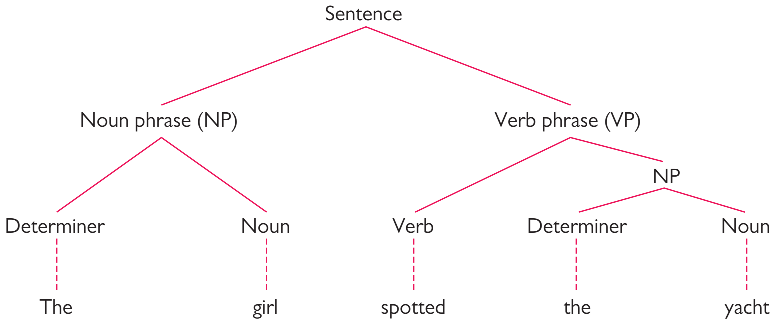


FIGURE 6.8 A simple phrase structure tree.

The constraints of our vocal and auditory systems dictate that words are uttered one by one in a serial fashion. However, according to many syntactic theories, this serial transmission obscures what is actually a hierarchical structure. Figure 6.8 illustrates the kind of syntactic structure that might be assigned to a simple sentence like 'The girl spotted the yacht'.

In this hierarchical analysis, each word is assigned a syntactic role in the sentence. The broken lines mark the links between each word and its role. These constituents are then grouped into phrases according to **phrase structure rules**, which are grammatical rules of English that indicate how phrases can combine. At the highest level the phrases combine to form a sentence. The analysis of sentences using **phrase structure grammar** offers a purely linguistic description, but we can see how it might apply to human language processing. On the perceptual side, we might think parsing should involve taking each word in a sentence, fleshing out its grammatical role and building a phrase structure tree that fits the sentence. This would not in itself recover the meaning of the sentence, but it would assist in this process by facilitating the **thematic role assignment** for the sentence (i.e. identifying the fact that the girl is the doer, that spotting is the activity she is doing, and that the yacht is what she is doing it to!).

Parsing is made more difficult by the fact that, as mentioned earlier, many words can be used in different grammatical roles. For example, the word *spotted* is used as a verb in Figure 6.8, but can be used as an adjective, as in a *spotted dress*. Equally, the noun *yacht* can be used as a verb, as in *to yacht around the world*. In the example sentence these alternative roles can be ruled out, because they would not form a grammatically coherent sentence in their alternative roles, but in many cases full sentences can be interpreted in different grammatically well-formed ways. Altmann (1998) gives the example of the sentence 'Time flies like an arrow', which has more than 50 grammatically permissible interpretations. For example, *time* appears to be acting as a noun but it could also be used as a verb, as in to 'time an egg'. Similarly, *flies* is most obviously a verb, but it could act as a plural noun – someone could time some flies! Semantically, such an interpretation may make little sense, but it could still be grammatical – just like Chomsky's famous example (Sentence 4 in Activity 6.3). In these cases we need to make use of more than just syntactic knowledge to resolve the ambiguity. The next section discusses how different models of parsing cope with ambiguities of this type. All models assume that we need to make use of multiple sources of information, but they differ in terms of the priority of the different information types.

4.2 Models of parsing

We found in Section 2.1 that the language system makes good use of the short time it takes to say a word. As speech enters the perceptual system, the cohort of potential candidates is whittled down, ensuring minimal delay in retrieving a word's meaning. One can ask the same question at the sentence level: does parsing assign a syntactic structure only at major syntactic boundaries (or even at the end of a sentence), or does it do so *incrementally*, refining the set of plausible syntactic structures every time a new word is recognized? It will not surprise you to learn that current models of sentence processing assume that parsing is incremental, and again this makes sense in terms of maximizing the availability of information for responding to the sentence. There are numerous demonstrations of incremental processing, employing a wide range of methods – an early example is the study of Tyler and Marslen-Wilson (1977). They made use of ambiguous phrases such as *landing planes*. With a preceding context such as 'If you walk too near the runway, . . .' the natural interpretation of *landing* is as an adjective (e.g. 'landing planes are dangerous' would be a suitable continuation), whereas following 'If you've been trained as a pilot, . . .' the interpretation is more likely to be as a verb (e.g. 'landing planes is easy'). Tyler and Marslen-Wilson wanted to know whether listeners showed a contextual bias in their parsing of the ambiguous phrase. If parsing is delayed until a syntactic boundary is reached, then there should have been no effect of preceding context on listeners' expectations about whether the word following *landing planes* was either *is* or *are*. They gauged listeners' expectations by presenting spoken fragments such as 'If you've been trained as a pilot, landing planes . . .' to participants and asking them to name a visual target word (either *is* or *are* in this case). They found that the speed of a naming response depended on the preceding context of the ambiguous phrase. Appropriate continuations were named quickly, compared with inappropriate ones. This is incompatible with the 'delayed parsing' hypothesis, because such a model predicts no effect of appropriateness. Instead it fits in with the idea that a plausible parse of a sentence is built up incrementally, and this influences expectations about upcoming words.

One of the most influential models of parsing, often known as the garden path model (Frazier, 1979), assumes that parsing is incremental, so each word is allocated a syntactic role as soon as it is perceived. In cases where more than one syntactic structure is compatible with the sentence so far, the parser makes a decision about which alternative to pursue based on syntactic information alone. The 'garden path' element comes in because the model predicts that there will be cases where the parse chosen at a point of ambiguity is incorrect (so the listener is 'led down the garden path'). Later in the sentence this incorrect selection will become clear, causing some backtracking as an alternative interpretation is attempted. The idea of pursuing some hypothesis and then reaching a dead-end requiring re-analysis fits in with people's intuitions about how they interpret some sentences. A famous example of 'garden pathing' is the sentence: 'The horse raced past the barn fell' (Bever, 1970). As you read this sentence, you may have had problems integrating the final word. Some people think that maybe there is an 'and' missing between *barn* and *fell*, or that there is a comma missing between *past* and *the*. But there is an alternative interpretation, which is a reduced version of 'The horse that was raced past the barn fell'. According to the garden path model, this alternative is not chosen when the word *raced* is first perceived, leading to trouble with interpretation later in the sentence.

The garden path model makes use of a set of guiding principles that specify which parse should be selected in the case of syntactic ambiguities, and these principles involve only syntactic information. The details of these principles are not essential – it is more important to keep in mind that the garden path model assumes a *serial* parser that maintains only one potential parse of a sentence at a time, and has an *autonomous* component in that the initial evaluation of a word's role in a sentence is based only on syntactic factors. In direct contrast, constraint-based models (e.g. MacDonald *et al.*, 1994) assume that parsing is *parallel* and *interactive*. So rather than maintaining a single syntactic analysis, these models allow more than one potential parse of a sentence to be evaluated at the same time (just as the cohort model of word recognition evaluates numerous candidates for word identification). Constraint-based models are thought of as interactive because they eliminate the autonomous stage of parsing assumed by the garden path model. Instead other factors, such as frequency and semantic plausibility, can influence parsing immediately.

MacDonald *et al.*'s model also increases the involvement of the lexicon in the parsing process, by assuming that some information about how a word can combine with other words is stored in the lexicon. By this kind of account, parsing becomes a bit like fitting the pieces of a jigsaw together. Each piece contains information about a word, including the kinds of syntactic context the word could fit into, and parsing involves fitting all the pieces together so that the words form a coherent sentence.

The two models described here are by no means the only models of parsing that researchers currently consider, but they do mark out the kinds of properties that generate debate in this area, and they highlight the kinds of questions that we need to investigate through experimentation. First and foremost among these, we need to try to address the question of whether parsing is autonomous, or whether it makes use of non-syntactic sources of information stored in the lexicon.

4.3 Is parsing autonomous?

As we have seen, the garden path model makes the strong prediction that the initial syntactic analysis of a word is unaffected by factors such as the meaning of the preceding context or the meaning of the words. In essence, the model puts all aspects of semantics aside until a word has been assigned a syntactic role. Initial data on the resolution of syntactic ambiguity showed garden path effects fully consistent with the autonomous approach of Frazier's model. In addition, some experiments designed specifically to look for semantic influences on syntactic ambiguity resolution found none. Ferreira and Clifton (1986) investigated how readers interpret verbs in phrases such as 'The defendant examined …'. Before you read on, think about how you might continue this sentence fragment. There are two common roles that 'examined' can play in this context. It could simply be the main verb of the sentence, as in 'The defendant examined his hands'. But it could also form part of what is known as a **reduced relative clause**. A relative clause might be 'The defendant that was examined by the lawyer …', and the reduced form would simply be the same but with 'that was' eliminated. The garden path model states that the preferred structure when *examined* is encountered is the more straightforward main verb interpretation. So if the sentence continuation is in fact a reduced relative structure, the Frazier model predicts a garden path effect when the true structure of the sentence becomes clear. So when a reader encounters 'The defendant examined by the lawyer turned out to be unreliable' they should show evidence of processing difficulty. This is exactly what Ferreira and Clifton found, using the eye-tracking methodology – people tended to fixate on the region just after the ambiguity, suggesting that they were having trouble incorporating the new information into their initially selected parse of the sentence.

The critical question here was whether the meaning of the word preceding the ambiguous verb could affect the garden path effect. So Ferreira and Clifton compared sentences like the one above to sentences like 'The evidence examined by the lawyer turned out to be unreliable'. In this case *evidence* is inanimate, which reduces the plausibility of the main verb interpretation (i.e. it seems unlikely that the evidence would examine anything). Despite the semantic bias towards the alternative reading, the garden path effect remained (i.e. fixation times remained long). On the surface, this seems like sound support for the autonomy assumed by the garden path model.

However, Trueswell *et al.* (1994) noticed that some of the contexts used by Ferreira and Clifton were less constraining than the example above. It is difficult to imagine a situation in which evidence could examine something, but Trueswell *et al.* argued that this was not the case for about half the materials used in the original experiment (e.g. 'the car towed …' where *car* is inanimate, but still quite a plausible candidate for something that tows). They ran another eye-tracking experiment using a similar design but with more constraining semantic contexts, and found that these contexts could lessen or even eliminate the garden path effect. The results of this and other similar studies are important because they show that, in some circumstances, the parsing system can be strongly affected by the semantic plausibility of the various parses of the system. The garden path model could perhaps be saved if it is assumed that the autonomous parsing component lasts only a short time, and that other factors come into play soon afterwards, but this greatly weakens the predictive power of the model, because it becomes harder to distinguish from models that allow semantic factors to play a stronger role. However, it is worth remembering that Frazier's syntactic constraints are not rendered immaterial by the finding that parsing is influenced by semantic plausibility. Instead, syntactic constraints appear to operate in *combination* with other constraints, with the ultimate goal being to weigh up the likelihood of different parses of a sentence in cases of ambiguity.

4.4 Constraints on parsing

It seems that the parsing system can be influenced by quite a number of different factors when it encounters an ambiguity. When a sentence is spoken, there is often useful information in the rhythm of the sentence. Think about how you might say the sentence 'Jane hit the man with the hammer' in the cases where (a) the man has a hammer, or (b) Jane has a hammer. One way to distinguish between these two possibilities is by changing your speech rate mid-sentence, so that different sets of words are grouped together. Of course these changes will be exaggerated when the speaker is aware of the potential ambiguity, but even in normal speech, the speaker can reduce ambiguity with changes in pitch and timing, and the listener can make use of this information (Warren, 1996). At a very different level, information about how often words are used in different syntactic structures can also influence the parsing process. This factor can be seen at work in the earlier example from Bever (1970), 'The horse raced past the barn fell'. One of the reasons this sentence causes so many problems is because the verb race is rarely used as a **past participle** (i.e. as in 'the horse that was raced . . .'). Not all verbs have this strong bias; for example *released* has the opposite bias – it is more likely to be used as a past participle (e.g. 'The hostage was released') than as a past tense of a main verb (e.g. 'The government released a press statement'). Trueswell (1996) showed that this lexical frequency factor also influences the way in which sentences are parsed. People seem to be able to keep track of the ways in which words are used in different sentences, and apply this knowledge in cases of ambiguity.

Perhaps the most striking example of a contextual influence on syntactic processing is based on the use of visual information. Tanenhaus *et al.* (1995) wanted to know whether the visual context of a sentence would affect the interpretation of syntactic ambiguities. In order to test this, they sat participants at a table on which some objects like apples and towels were placed and gave them instructions to move the objects, such as 'Put the apple on the towel in the box'. The participants wore head-mounted eye trackers so that the experimenters could monitor eye movements as the sentences were heard (see Box 6.2 in Section 2.2.3). The sentences had a temporary syntactic ambiguity, which in the case of the example here involves the phrase 'on the towel'. We know from studies like Ferreira and Clifton (1986) that when people hear 'Put the apple on the towel . . .' they tend to interpret 'on the towel' as the desired destination of the apple. But the continuation '. . . in the box' should force a reassessment of the sentence (i.e. the sentence is a reduced form of 'Put the apple that's on the towel in the box'). We have seen that various sentential factors such as semantic plausibility can reduce or eliminate this garden path effect, but what about external, environmental context? Tanenhaus *et al.* (1995) gave the participants instructions in two types of external context (see Figure 6.9). In one case (see Figure 6.9(a)), there was an apple on a towel, another towel, and a box. This context supports the initial interpretation of 'on the towel' as referring to the destination, so people tended to look at the apple and then at the empty towel, and only looked at the true intended destination once the disambiguating speech ('in the box') was heard. However, when the scene also included a second apple on a napkin (see Figure 6.9(b)), participants' eye movements were quite different. Now when they heard 'on the towel', they rarely looked at the empty towel, because they interpreted 'on the towel' as distinguishing information between the two apples (one was on a towel and one on a napkin). In other words, the environmental situation provided a source of information that could eliminate the garden path effect.

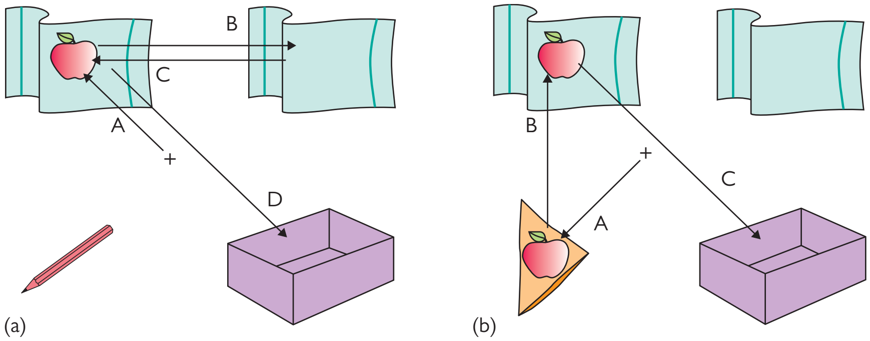


FIGURE 6.9 Two visual contexts from Tanenhaus *et al.* (1995) showing the typical sequence of eye movements in response to the ambiguous instruction 'Put the apple on the towel in the box'. Eye fixations began at the central cross, and continued in the sequence indicated by the capital letters. *Source: based on Tanenhaus et al., 1995, Figures 1 and 2*

SUMMARY OF SECTION 4

- Understanding a sentence requires a parsing process in which each word is assigned a grammatical role.
- The garden path model assumes that the parser operates autonomously, without any influence of non-syntactic factors.
- Recent studies of syntactic ambiguity resolution suggest that a variety of different constraints can influence parsing, including even the environment of the listener.

5 CONCLUSION

This chapter has provided a brief account of some of the main components of the language system, particularly with reference to recognizing words and understanding sentences. We have seen that many of the processes involved can be modelled in terms of competition between multiple candidates, implying that the language system is busy evaluating countless hypotheses about an utterance at numerous levels at any moment in time. Thankfully we remain blissfully unaware of these operations, with only a pretty terse 'executive summary' of the process available to conscious awareness.

Another recurring theme has involved the extent to which components of the system operate independently of each other. There is a long way to go in this debate, but the current state of play seems to be one in which there is a surprising level of linkage between sub-systems. So reading a word engages processes and representations related to speech perception, and the way in which you process a spoken sentence can be influenced by the real world context in which you hear it. This interconnectedness may well reflect two aspects of language processing: the complexity of language, and the speed with which we need to communicate. In terms of language development, it makes a lot of sense to reuse existing mechanisms when we are trying to add a new mechanism such as the mechanism for reading. In terms of adult language processing, it makes sense to call on as much useful information as possible to minimize the time it takes to comprehend a sentence.

Answer to Activity 6.1

The approximate word boundary positions are marked in Figure 6.10, along with the words themselves. Some

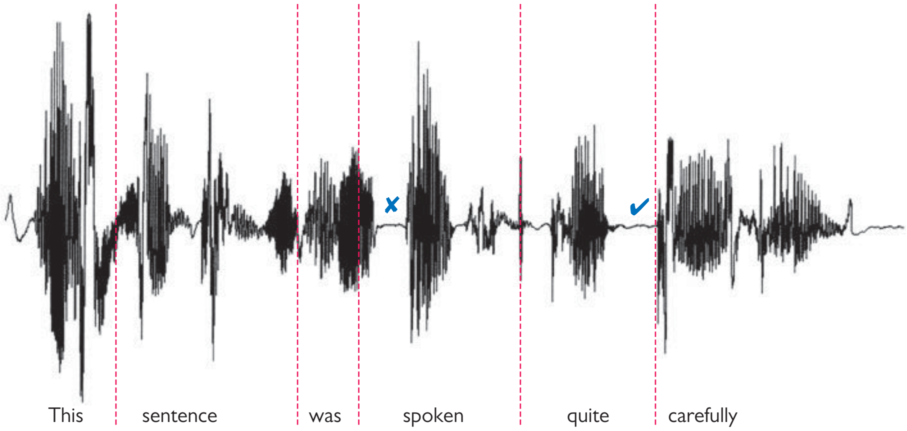


FIGURE 6.10 Typical word boundaries in a fluent sentence.

gaps in the speech (low amplitude signal) are aligned with word boundaries (e.g. between *quite* and *carefully*, marked with a tick), whereas others are not (e.g. within *spoken*, marked with a cross). In general, short periods of silence are poor indicators of word boundaries, meaning that we have to find better ways to segment speech.

FURTHER READING

Altmann, G.T.M. (2001) 'The language machine: psycholinguistics in review', *British Journal of Psychology*, vol.92, pp.129–70. A comprehensive review of the field of psycholinguistics.

Harley, T. (2008) *The Psychology of Language: From Data to Theory*, Hove, Psychology Press. An up-to-date and readable overview of language processing. Topics covered include spoken and visual word recognition and sentence comprehension, as well as additional topics such as first and second language acquisition and language production that are not covered in this chapter.

McQueen, J. (1996) 'Word spotting', *Language and Cognitive Processes*, vol.11, no.6, pp.695–9. A very accessible overview of the wordspotting task, highlighting key studies that have used this task, as well as discussing the advantages and limitations of using wordspotting to study speech segmentation and lexical access.

REFERENCES

Altmann, G.T.M. (1998) 'Ambiguity in sentence processing', *Trends in Cognitive Sciences*, vol.2, no.4, pp.146–52.

Berko, J. (1958) 'The child's learning of English morphology', *Word*, vol.14, no.2, pp.150–77.

Bever, T.G. (1970) 'The cognitive basis for linguistic structures', in Hayes, J. (ed.) *Cognition and the Development of Language*, New York, Wiley.

Bowers, J.S., Davis, C.J., Mattys, S.L., Damian, M.F., and Hanley, D. (2009) 'The activation of embedded words in spoken word identification is robust but constrained: evidence from the picture-word interference paradigm', *Journal of Experimental Psychology: Human Perception and Performance*, vol.35, no.5, pp.1585–97.

Chéreau, C., Gaskell, M.G., and Dumay, N. (2007) 'Reading spoken words: orthographic effects in auditory priming', *Cognition*, vol.102, pp.341–60.

Collins, A.M. and Loftus, E.F. (1975) 'A spreading-activation theory of semantic processing', *Psychological Review*, vol.82, no.6, pp.407–28.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001) 'DRC: a dual route cascaded model of visual word recognition and reading aloud', *Psychological Review*, vol.108, no.1, pp.204–56.

Cunillera, T., Càmara, E., Laine, M., and Rodríguez-Fornells, A. (2010) 'Words as anchors: known words facilitate statistical learning', *Experimental Psychology*, vol.57, no.2, pp.134–41.

Cutler, A. and Carter, D.M. (1987) 'The predominance of strong initial syllables in the English vocabulary', *Computer Speech and Language*, vol.2, no.2, pp.133–42.

Cutler, A. and Norris, D. (1988) 'The role of strong syllables in segmentation for lexical access', *Journal of Experimental Psychology, Human Perception and Performance*, vol.14, no.1, pp.113–21.

Cutler, A. and Otake, T. (2002) 'Rhythmic categories in spoken-word recognition', *Journal of Memory and Language*, vol.46, no.2, pp.296–322.

Ferreira, F. and Clifton, C. (1986) 'The independence of syntactic processing', *Journal of Memory and Language*, vol.25, no.3, pp.348–68.

Frazier, L. (1979) *On Comprehending Sentences: Syntactic Parsing Strategies*, Bloomington, IN, Indiana University Linguistics Club.

Gaskell, M.G. and Marslen-Wilson, W.D. (2002) 'Representation and competition in the perception of spoken words', *Cognitive Psychology*, vol.45, no.2, pp.220–66.

Glushko, R.J. (1979) 'The organization and activation of orthographic knowledge in reading aloud', *Journal of Experimental Psychology: Human Perception and Performance*, vol.5, no.4, pp.674–91.

Grainger, J. and Jacobs, A.M. (1994) 'A dual read-out model of word context effects in letter perception – further investigations of the word superiority effect', *Journal of Experimental Psychology: Human Perception and Performance*, vol.20, no.6, pp.1158–76.

Hauser, M.D., Newport, E.L., and Aslin, R.N. (2001) 'Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins', *Cognition*, vol.78, no.3, B53–B64.

Kellenbach, M.L., Wijers, A.A., and Mulder, G. (2000) 'Visual semantic features are activated during the processing of concrete words: event-related potential evidence for perceptual semantic priming', *Cognitive Brain Research*, vol.10, no.1–2, pp.67–75.

Kutas, M. and Hillyard, S.A. (1980) 'Reading senseless sentences: brain potentials reflect semantic incongruity', *Science*, vol.207, no.4427, pp.203–5.

Lucas, M. (1999) 'Context effects in lexical access: a meta-analysis', *Memory and Cognition*, vol.27, no.3, pp.385–98.

Lucas, M. (2000) 'Semantic priming without association: a meta-analytic review', *Psychonomic Bulletin and Review*, vol.7, no.4, pp.618–30.

MacDonald, M.C., Pearlmutter, N.J., and Seidenberg, M.S. (1994) 'Lexical nature of syntactic ambiguity resolution', *Psychological Review*, vol.101, no.4, pp.676–703.

Marslen-Wilson, W.D. (1987) 'Functional parallelism in spoken word recognition', *Cognition*, vol.25, no.1, pp.71–102.

Marslen-Wilson, W.D. and Welsh, A. (1978) 'Processing interactions and lexical access during word recognition in continuous speech', *Cognitive Psychology*, vol.10, no.1, pp.29–63.

Marslen-Wilson, W.D., Tyler, L.K., Waksler, R., and Older, L. (1994) 'Morphology and meaning in the English mental lexicon', *Psychological Review*, vol.101, no.1, pp.3–33.

Masson, M.E.J. (1995) 'A distributed memory model of semantic priming', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.21, no.1, pp.3–23.

McClelland, J.L. and Elman, J.L. (1986) 'The TRACE model of speech perception', *Cognitive Psychology*, vol.18, no.1, pp.1–86.

McClelland, J.L. and Rumelhart, D.E. (1981) 'An interactive activation model of context effects in letter perception. Part 1: an account of basic findings', *Psychological Review*, vol.88, no.3, pp.375–407.

McConkie, G.W. and Rayner, K. (1975) 'The span of the effective stimulus during a fixation in reading', *Perception and Psychophysics*, vol.17, no.6, pp.578–86.

McQueen, J.M., Norris, D., and Cutler, A. (1994) 'Competition in spoken word recognition – spotting words in other words', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.20, no.3, pp.621–38.

Nelson, D.L., McEvoy, C.L., and Schreiber, T.A. (1998) *The University of South Florida Word Association, Rhyme, and Word Fragment Norm* [online]. Available from http://w3.usf.edu/FreeAssociation (accessed 19 January 2004).

Norris, D., McQueen, J.M., and Cutler, A. (2000) 'Merging information in speech recognition: feedback is never necessary', *Behavioral and Brain Sciences*, vol.23, no.3, pp.299–370.

O'Regan, J.K. and Jacobs, A.M. (1992) 'Optimal viewing position effect in word recognition – a challenge to current theory', *Journal of Experimental Psychology: Human Perception and Performance*, vol.18, no.1, pp.185–97.

Rastle, K., Davis, M.H., and New, B. (2004) 'The broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition', *Psychonomic Bulletin & Review*, vol.11, no.6, pp.1090–8.

Rayner, K. and Duffy, S.A. (1986) 'Lexical complexity and fixation times in reading – effects of word-frequency, verb complexity, and lexical ambiguity', *Memory and Cognition*, vol.14, no.3, pp.191–201.

Rodd, J., Gaskell, G., and Marslen-Wilson, W. (2002) 'Making sense of semantic ambiguity: semantic competition in lexical access', *Journal of Memory and Language*, vol.46, no.2, pp.245–66.

Saffran, J.R., Aslin, R.N., and Newport, E.L. (1996) 'Statistical learning by 8-month-old infants', *Science*, vol.274, no.5294, pp.1926–8.

Saffran, J.R., Johnson, E.K., Aslin, R.N., and Newport, E.L. (1999) 'Statistical learning of tone sequences by human infants and adults', *Cognition*, vol.70, no.1, pp.27–52.

Seidenberg, M.S. and McClelland, J.L. (1989) 'A distributed, developmental model of word recognition and naming', *Psychological Review*, vol.96, no.3, pp.523–68.

Seidenberg, M.S. and Tanenhaus, M.K. (1979) 'Orthographic effects on rhyme monitoring', *Journal of Experimental Psychology: Human Learning and Memory*, vol.5, pp.546–54.

Shillcock, R., Ellison, T.M., and Monaghan, P. (2000) 'Eye-fixation behavior, lexical storage, and visual word recognition in a split processing model', *Psychological Review*, vol.107, no.4, pp.824–51.

Starr, M. and Rayner, K. (2001) 'Eye movements during reading: some current controversies', *Trends in Cognitive Sciences*, vol.5, no.4, pp.156–63.

Swinney, D.A. (1979) 'Lexical access during sentence comprehension: (re)consideration of context effects', *Journal of Verbal Learning and Verbal Behavior*, vol.18, no.5, pp.645–59.

Tabossi, P. and Zardon, F. (1993) 'Processing ambiguous words in context', *Journal of Memory and Language*, vol.32, no.3, pp.359–72.

Taft, M. and Forster, K.I. (1975) 'Lexical storage and retrieval of prefixed words', *Journal of Verbal Learning and Verbal Behavior*, vol.14, no.6, pp.638–47.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., and Sedivy, J.C. (1995) 'Integration of visual and linguistic information in spoken language comprehension', *Science*, vol.268, no.5217, pp.1632–4.

Trueswell, J.C. (1996) 'The role of lexical frequency in syntactic ambiguity resolution', *Journal of Memory and Language*, vol.35, no.4, pp.566–85.

Trueswell, J.C., Tanenhaus, M.K., and Garnsey, S.M. (1994) 'Semantic influences on parsing – use of thematic role information in syntactic ambiguity resolution', *Journal of Memory and Language*, vol.33, no.3, pp.285–318.

Tyler, L.K. and Marslen-Wilson, W.D. (1977) 'The on-line effects of semantic context on syntactic processing', *Journal of Verbal Learning and Verbal Behavior*, vol.16, no.6, pp.683–92.

Van Orden, G.C. (1987) 'A rows is a rose – spelling, sound, and reading', *Memory and Cognition*, vol.15, no.3, pp.181–98.

Van Petten, C. and Kutas, M. (1987) 'Ambiguous words in context – an event-related potential analysis of the time course of meaning activation', *Journal of Memory and Language*, vol.26, no.2, pp.188–208.

Warren, P. (1996) 'Prosody and parsing: an introduction', *Language and Cognitive Processes*, vol.11, no.1–2, pp.1–16.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

CHAPTER 7

LANGUAGE IN ACTION

Simon Garrod and Tony Sanford

1 INTRODUCTION

In the previous chapter we saw how language comprehension can be viewed as a translation process. The process starts with patterns of sounds or letters and then translates them into increasingly large and more abstract units – words, phrases, and sentences – until the reader or listener eventually arrives at the meaning. However, this view of language processing does not take into account how language is used in everyday settings and how this affects processing. Here we go beyond the model of the previous chapter to look at language in action and how language comprehension and language production reflect the different circumstances under which the language is being used.

Language is used primarily for verbal communication and most often through speech rather than writing. From the moment you get up in the morning until you finally fall asleep at night you will be speaking to someone or other almost every minute of the day. Furthermore, when you do this you are nearly always engaging in a dialogue, in which one or more people interact directly with you and with each other. Of course, you can also use written language to communicate. Most often it is through monologue, in which there is no interaction between the writer and the reader, as when you write an essay or read a newspaper article. But increasingly even written communication is becoming interactive. Consider, for instance, exchanging text messages with a friend or taking part in a live 'chat' conversation on your computer. Whether it is through speech or writing, communication is perhaps the most important social, cultural, and cognitive activity that we engage in, so understanding how we use language to communicate is central to the study of human cognition.

Although it can easily be argued that spoken language and in particular dialogue is the more natural form of language use, we begin this chapter with a look at research on written language, and only then proceed to dialogue. This sequence reflects the history of experimental research on the psychology of language. Much more research has been done on written language comprehension than on dialogue, and the findings of this research illustrate important features of language in action that take us beyond the simple translation metaphor discussed above.

One way in which written language comprehension takes us beyond the translation metaphor is that it requires the integration of information across different sentences in a text. As we see below, a key aspect of this integration is that it depends upon access to non-linguistic background knowledge. To this extent language processing in this broader sense is less 'encapsulated' than it seemed in the previous chapter. The precise interpretation of any real piece of language calls upon a wide range of different sources of information, including our knowledge of the situation under discussion.

We begin by defining what is meant by **text** and then consider what this means for text processing.

2 WRITTEN LANGUAGE AND DISCOURSE

Traditionally linguists have identified two characteristics that differentiate a text from just a collection of isolated sentences. The first is what they call cohesion, and the second is known as **coherence** (Halliday and Hasan, 1976; Brown and Yule, 1983).

Texts are cohesive to the extent that they contain many expressions whose interpretation depends in some way on interpretations of prior expressions in the text, and these co-interpretations serve to link the sentences together. One major source of cohesion comes from repeated reference or anaphora. For instance, the following pair of sentences are cohesive because the pronouns *she* and *it* in (2) take their interpretation from the noun-phrases *Susan* and *some money* in (1):

(1) Bill wanted to lend Susan some money.

(2) She was hard up and really needed it.

Furthermore, the cohesive link contributes to the fact that sentences (1) and (2) constitute a piece of text. But cohesion is not all there is to bind sentences together into a text. For instance, consider the following variant of sentences (1) and (2):

(1') Bill wanted to lend Susan some money.

(2') It is not nice to have close friends who are really hard up.

Here there are no cohesive anaphoric links between the sentences, yet we still have an acceptable text. What is important in this case is that the two sentences can be related into a coherent whole through inference. The reader will take it that the unpleasantness of having friends who are hard up is the reason why Bill wants to lend Susan some money, and by implication that Susan is Bill's friend. So a text's coherence comes from establishing the logical and psychological consistency between the events and states portrayed.

Cohesion and coherence are not independent. Even in texts such as (1'–2') above, there is a kind of cohesive bond set up because it is assumed that 'Susan' must be an instance of one of 'Bill's close friends'. In fact it will often be the case that the interpretation of cohesion markers, such as pronouns, depends upon establishing coherence, and vice versa. For instance, consider the following further variant of sentences (1) and (2):

(1") Bill wanted to lend his friend some money.

(2") He was hard up and really needed it.

and

(2"') However, he was hard up and couldn't afford to.

The same pronoun *he* in almost identical clauses (2" and 2"') takes on different referential interpretations depending upon the different coherence relations between the two sentences. At the same time the form of coherence relation differs depending on the assignment of the pronoun. For instance, while his being hard up in (2") is taken as a reason for Bill's wanting to lend the money, his being hard up in (2‴) is taken as an obstacle to Bill lending the money.

So, collections of sentences become texts through the links that bind them together into a coherent structure. Some of the links are explicitly signalled through cohesion markers, such as pronouns or sentence connectives like *but*, *therefore*, and *however*, whereas other links depend upon inferring the logical or psychological relationships between the events portrayed. Besides reference, there are many other sources of linkage. For instance, in narrative text there have to be temporal links that order the events in the story. In simple cases these are signalled with explicit temporal expressions as in the following short passage:

> *(3) Yesterday Mary visited (e_1) her grandmother. Later, she stopped (e_2) at a shop to buy some flowers. (e_1 and e_2 denote events)*

Here the events are explicitly ordered through the temporal cohesion device *later*. So event e_1 precedes event e_2. But again, ordering often comes from establishing a coherent chain of events. For instance, in the following variant the 'visiting' and the 'stopping at the shop' are interpreted as occurring in the opposite order:

> *(4) Yesterday Mary visited (e_2) her grandmother. She stopped (e_1) at a shop to buy some flowers. She then went and presented them to her as a gift for her eightieth birthday.*

So temporal cohesion, like referential cohesion, often depends upon the coherence of the passage as a whole.

Examples like those above where there is no explicit marker to indicate how the sentences relate to each other show that the coherence of a discourse is in the mind of the reader. In all of the examples above the reader uses general knowledge to make a coherent connection. Brown and Yule (1983) contrast the Discourse-as-Product view (the coherence is in the text alone) with Discourse-as-Process, where the coherence comes from mental processes called upon to interpret the text. The second view leads us naturally to psychological investigations of interpretation, and how the mind adds to what is in the text.

2.1 Processes underlying text interpretation

Much research on language comprehension is concerned with how the comprehension process operates in real time: when each of the sub-operations occur that eventually lead to a coherent understanding of the text. Here we shall concentrate on just four issues, each of which concerns the establishment of coherence, and extracting the meaning of discourse beyond just the meaning of the words it contains.

2.1.1 Anaphora resolution

As we have already seen, anaphora is crucial to text cohesion. So it is not surprising that how we resolve anaphors during comprehension is one of the most-studied components of text comprehension (Garnham, 2001). When reading a text, it is important to keep track of who is doing what and when. For instance, given the passage in Box 7.1, we need to know what *John* did and what *William* did.

Right until the last sentence, every time we encounter *he*, we interpret it as standing for *John* (it is an anaphor for *John*). *He* does not stand for *Mary* because *Mary* is female, and *he* specifies a singular and male antecedent. There is no ambiguity, and we might suppose that the language processor checks to see what *he* stands for as soon as *he* is encountered. *John* is the

BOX 7.1 A sample of coherent textt

We shall use this example to discuss a number of processing issues throughout this chapter.

John went to the shops with Mary. She went off to buy clothes, and he went to the bank. With the money he had, he was going to buy some new CDs. On the way to the shop, he bumped into William. He found out William hadn't had any lunch that day. John lent William some money because he was hard up.

main character in the story, being mentioned often, and so even when a new male individual turns up, there is a preference to equate *he* with *John*.

In the last sentence of the text in Box 7.1, something different happens. Here *he* stands for *William*. The new pronoun assignment works because the reader can use his or her knowledge to disambiguate the pronoun. As we saw earlier, the reader can assume that people who have money are in a position to lend, and people who are hard up need money. So, putting these facts together with what the text says enables the processor to resolve *he* as referring to *William*.

In this hypothetical analysis, we can see several potential sources of information that the comprehension system might use to resolve pronoun-based (pronominal) anaphora:

- Gender cues in the pronouns. Apart from gender marking (*he* versus *she*), we also have animacy marking (*he, she* vs *it*) and reflexivity (*he* versus *himself*, *she* versus *herself*).
- Main character versus secondary character (or continued reference to one character).
- General knowledge, e.g. people with X can give X, while people without can't. People without X might want X, etc. These properties form a very large set of general knowledge beliefs.

These are just some of the cues that are known to support anaphora resolution. A full discussion of anaphora from a psychological standpoint is given in Garnham (2001).

2.1.2 When word-meaning is used

The text in Box 7.1 illustrates other problems that psycholinguists have worked on intensively. One is the question of when word meaning enters into the comprehension process. A useful technique for studying when meaning may be accessed is to track a person's eye-movements during reading, as described in Chapter 6. A good way of investigating when word meaning is accessed is to insert into the text a word that is not appropriate, or anomalous, and see if it disturbs the pattern of eye-movements as soon as it is encountered. If it does (causing longer initial fixations, or causing an increase in regressive eye-movements), then one can conclude that the word's meaning is being used **at that point**.

Traxler and Pickering (1996) compared how people read materials like (5) and (6) below:

> (5) *That's the pistol with which the man shot the gangster yesterday afternoon.*
>
> (6) *That's the garage with which the man shot the gangster yesterday afternoon.*

The word *shot* fits sentence (5) in meaning, but not sentence (6). Participants read sentences of this type while having their eye-movements monitored. Traxler and Pickering found that the very first fixations on the word *shot* were longer in the implausible cases, like (6), than in the plausible cases, like (5). So, the meaning of *shot* must have been accessed, and incorporated into the meaning of the sentences as a whole, as soon as the word was fixated. Results such as these are consistent with what is called **incremental interpretation**, a view of discourse comprehension that says each word is interpreted and incorporated into the meaning of the sentence as soon as it is encountered. See Chapter 6 for related work that investigates accessing word meaning.

2.1.3 Non-literal meaning

The immediacy of processing observed in the studies above is consistent with what one might term the 'standard view' of text understanding. In this view, as words are encountered, their meanings are retrieved from long-term (semantic) memory. As the sentence unfolds, the syntactic structure is derived, and the meanings of the words are then combined to give a sentence meaning. However, there are several problems with this view. One is the problem of non-literal meaning. Consider the sentence, *John asked the man if he could tell him the time.* Although *Can you tell me the time?* is a literal question (to which the answer is 'yes' or 'no'), the interpretation given is as a request (*Tell me what the time is, please*). These are known as *indirect speech acts*. How do people get the correct interpretation? The standard account

(Searle, 1980) has the following steps (Glucksberg and Keysar, 1990):

- Derive a literal interpretation.
- Assess that interpretation against the context of the utterance.
- If (and only if) literal meaning is a poor fit, derive a non-literal interpretation.

This suggests that indirect speech acts should take longer to process than direct speech acts (e.g. questions that are in fact questions and not indirect requests). Certainly for some cases the model is wrong, because the indirect cases are processed just as quickly as the direct cases (e.g. Gibbs, 1983). Now look at the following sentence from Box 7.1: *On the way to the shop, he bumped into William*. This is interpreted as John meeting William, probably unexpectedly, and not actually colliding with him. It requires non-literal interpretation. Now according to the model above, a statement will only get a non-literal interpretation if it is needed. Glucksberg *et al.* (1982) showed that non-literal interpretations may be given when they are *not* needed, suggesting that they are automatically derived. They asked participants to decide whether statements were **literally** true, for instance, *Some desks are junkyards*. What they showed was that the time to answer 'no' was longer in such cases than in the case of literally false statements with no metaphoric interpretation, such as *Some desks are roads*. They suggest that this is because the non-literal meaning is highly available from memory, and so it intrudes in the 'true' judgement, leading to longer times to decide that it is not literally true.

It might be thought that such cases are rare, but they are not. In particular, Lakoff (1987) has provided numerous examples of metaphors used in everyday language. Here are a few:

(7) When John heard about his wife, he exploded.

(8) When it came to chemistry, Fred was a little rusty.

(9) It is morally right to fight poverty.

Thinking about the literal meaning of these sentences quickly shows that they are metaphorical. It is easy to think of your own examples of such things, and they provide a challenge for any view that says sentence meaning is based on the literal meanings of words.

2.1.4 Inferences

We have seen that general knowledge is needed to understand texts. The processes that give rise to coherence are known as inference-making. In psycholinguistics, a distinction has been drawn between necessary inferences and elaborative inferences. Here is a case where a necessary inference allows interpretation:

(10) Mary got some picnic things out of the car. The beer was warm.

(11) Mary got some beer out of the car. The beer was warm.

The inference in (10) is that the beer is part of the picnic things, and so came out of the car. Haviland and Clark (1974) were the first to show that such inferences take time. Using self-paced reading, they showed that the reading time for the *second* sentence was longer by some 100 ms for example (10), where *beer* hadn't been mentioned before, than for (11), where it had. The extra time is the time to form the inference *beer is part of the picnic things*, which forms a link, or bridge, and is sometimes called a **bridging inference**. Unless the inference had been drawn, there would be no way in which the two sentences could be sensibly connected – it would not be a coherent discourse. That is why such inferences are called necessary.

Elaborative inferences refer to inferences that are not strictly necessary. Consider (12) below:

(12) Unable to control his rage, the angry husband threw the valuable porcelain vase at the wall.

Did you make the inference that the vase broke? It is certainly a plausible inference, as are several others, like *he was having a row with his wife*, for instance. Such inferences are not necessary for understanding the sentence. In fact they are defeasible (can be cancelled). For instance, if you read *He missed and the vase landed on the sofa* you would have to cancel your inference that the vase broke. Research into whether elaborative inferences are made has been carried out by using a variety of priming techniques, such as showing the word 'broke' after (12). If it is primed in a lexical decision task, then that would suggest that the inference had been made. The evidence suggests that elaborative inferences are not made routinely (McKoon and Ratcliff, 1992), but it remains to be shown exactly what the conditions are under which they are and are not made.

2.1.5 Relating language to knowledge

Almost all of the processes discussed above show how world-knowledge is needed to interpret the significance of what is being said in a text. This leads to a variety of questions concerning the text-knowledge interface. One very important concept is that understanding depends upon the reader setting up a mental model of what the text is depicting. Look at texts (13) and (14):

> *(13) Harry put the wallpaper on the table. Then he sat his cup on it.*
>
> *(14) Harry put the wallpaper on the wall. Then he sat his cup on it.*

Example (14) sounds odd because when wallpaper is on the wall it is in a vertical plane and would not support a cup. If you noticed this, which you will have, then this means you produced a mental model of what the first sentence meant in relation to the real world, and when you integrated the second sentence, there was a problem. Approaches to language understanding all include some reference to how world-knowledge is involved in interpretation (e.g. Kintsch, 1988; Sanford and Garrod, 1998). Here we want to emphasize that it is very important for readers to have the right mental model if they are to understand the discourse. Sanford and Garrod (1981, 1998) believe that much of our knowledge is organized in situation-specific packages. For instance, if one reads about buying something in a shop, the important aspects of what this entails become available as part of one's mental representation of what the text is about. This means, for instance, that the writer can refer to things that have not been explicitly mentioned before, as in (15):

> *(15) The court case was going badly for the defendant. He could see that the judge had no time for him.*

Encountering *the judge* for the first time causes no difficulty, taking no more time to process than if the judge had been mentioned explicitly (see Sanford and Garrod, 1998, for further details). These examples show how language accesses information in memory that represents situations and settings, containing information about what to expect to be present at a court case, for instance. Precisely how situational information is represented has been a considerable area of enquiry. Recently, one intriguing idea is that understanding relies on representations that are literally of how our bodies interact with the world. This is quickly becoming a key issue in how understanding works, and what meaning really is.

2.1.6 Knowledge, meaning, and embodiment

In the traditional view of language processing, concepts are treated in an abstract way. Indeed, the meanings of words are commonly thought of as being represented as mental lists or networks of attributes (see Section 3 on the mental lexicon in Chapter 6). According to this view, language conveys meaning by using abstract words, combined by syntactic rules (e.g. Fodor, 2000; Kintsch, 1988; Pinker, 1994). However, an alternative view has been emerging, in which meaning is rooted in perception and bodily action – literally, in how we interact with the world.

One motivation behind this embodiment view of meaning comes from what has been called the **symbol grounding problem**. Harnad (1990) illustrates the problem nicely:

> *Suppose a person P lands at an airport in a country where they do not speak the language. P has a dictionary written solely in that language. P sees a sign at the airport written in the local script, and looks up the sign in the dictionary to see what it means. The first part of the sign has an entry described in terms of more of the same script, so P has to look up what the first part of that description means, in the same dictionary. No matter how many times P does this, the meaning of the original will not be found! The*

> *dictionary does not, on its own, contain the meaning of the expressions on the sign.*

So, according to Harnad (see also Searle, 1980) symbols can only have meaning by being related to things in the world, and not to other symbols and words. Consider the words *left* and *right*. Definitions for these in dictionaries make interesting reading. One element with *left*, for instance, is that it is the opposite of *right*. Without some means of interpreting that statement, it simply doesn't make sense. Typically in dictionaries, there is reference to the outside world. So, *left* is defined as 'that side in which a person has normally the weaker and less skilled hand', and right as 'that side in which a person has normally the stronger and more skilled hand'. Unless one knows about people, and which side (regardless of name) is normally stronger, the definitions are vacuous. So, the meanings of words have to be grounded in the world. Activity 7.1 considers another case, the verb *trudge*.

The view that cognitive activities such as understanding the meaning of something are bound up with representations of actual interactions with the world is part of the issue of **embodied cognition**. It is a short step from words to sentences. Sentences suffer from the symbol grounding problem if they are not connected to perception and action. For instance, in order to understand the unconventional use of the word *elbow* in verb form (e.g. *John elbowed the pencil to Mary*), we have to make use of the range of actions of which the elbow is capable.

Is there any experimental evidence for bodily involvement in the understanding of entire *sentences* depicting simple actions? Recent work by Glenberg and Kaschak (2002) and colleagues suggests that the direction of real movement underpins the comprehension of transfer and movement sentences. Glenberg and Kaschak (2002) compared people's responses to two types of sentence, examples of which included imperative, and physical and abstract transfer sentences:

TOWARD SENTENCES

Open the drawer (IMPERATIVE)

Courtney handed you the notebook (TRANSFER: DOUBLE OBJECT)

Liz told you a story (ABSTRACT TRANSFER)

AWAY SENTENCES

Close the drawer (IMPERATIVE)

You handed Courtney the notebook (TRANSFER: DOUBLE OBJECT)
You told Liz a story (ABSTRACT TRANSFER)

Physical **towards** sentences require a movement towards the body to be executed, while **away** sentences require a movement away. Participants were presented with sentences of these two types, along with nonsense sentences like *Boil the air*. The task was to judge whether each sentence was sensible. They did this for each sentence by pressing a button that was on the far end of a box (away from the body), or near to the body. For half of the participants, 'Yes' was the near button, while for the other half, it was far. At the outset of each trial, the response finger rested on a centre button, so that to respond participants actually had to move either towards themselves or away.

The prediction was that when 'away' sentences were comprehended, part of the understanding involves a

✱ ACTIVITY 7.1

Try to write down a definition of *trudge* before moving on.

Try to write down a definition of *waltz*.

A word-description of *trudge* is virtually impossible, as well as meaningless. This is because *trudge* defines a set of motion attributes that are embodied in (human) movement. This is obvious with *waltz* too. So, one way of trying to get around the symbol grounding problem is to assume that meanings relate to representations of our physical interactions with the world.

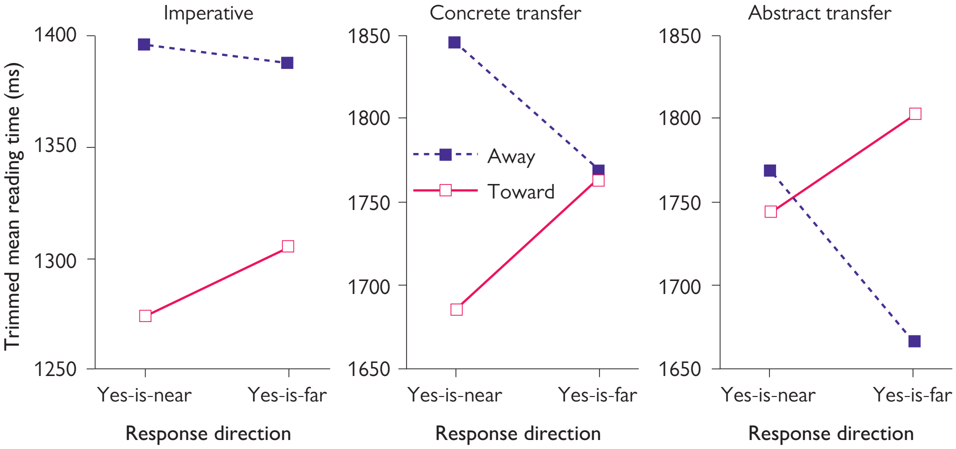


FIGURE 7.1 Results of the experiment by Glenberg and Kasheck. *Source: Glenberg and Kaschack (2002)*

simulation of transferring the object (concrete or abstract) away from the body, so responses that are physically 'away' should be made more rapidly than responses that are 'towards' movements. The opposite should apply to 'towards' sentences. The results in Figure 7.1 show that in the yes-is-near condition, *away* sentences took longer to process than *towards*, while the opposite was true for *towards* sentences. Although the effects appear weaker for the imperative sentences, the difference from the other conditions was not reliable. Glenberg and Kasheck take these results as supporting their view that the simple understanding of transfer sentences is rooted in the actions underlying the transfers themselves.

2.2 Special topics

Before moving on to language production and dialogue, we shall focus on two relatively new, emerging, and important issues: shallow processing and perspective in communication.

2.2.1 Incomplete processing and selective processing

Just how completely do we utilize the meanings of words in establishing sentence meaning? Just how much detail goes into the representations of discourse we end up with after reading? Quickly complete the questionnaire in Activity 7.2. (Do this now before reading on.)

The failure to use the full meaning of a word is a demonstration of shallow processing: not shallow in the sense of being sensory rather than to do with meaning, but rather in the sense of dealing only superficially with meaning. This can easily be seen with contrasts between different versions of the Moses illusion. For example, if *Adam* is substituted for *Moses*, then everyone spots that he didn't put any animals on the Ark (Van Oostendorp and De Mul, 1990). It is argued that this is because *Moses* is more similar in 'meaning' to *Noah* than *Adam* is (according to participants' ratings). So, it is not that people don't process the meanings of words in anomalies; rather, they do not process them very deeply. (See Barton and Sanford (1993) for further evidence, using the anomaly *After an aircrash, where should the survivors be buried?*)

In Chapter 6 and in Section 2.1.2 of this chapter, we saw evidence suggesting that word meaning is retrieved immediately when a word is read. Are the findings with anomalies inconsistent with the incremental interpretation hypothesis? No. First, if a word is a really poor fit in meaning, it is noticed; it is just when it is a close (but wrong) fit that there are problems. So, meanings may be used immediately, but only part of the full meaning may be used. For this reason, it may be best to say that readers immediately initiate meaning retrieval, but that this may be incomplete.

✱ ACTIVITY 7.2

Underspecification and depth of processing

Complete the following questionnaire as quickly as possible (all questions apply to what is true in the United Kingdom).

a. Can a 16-year-old girl marry without her parents' permission?
b. Can a man marry his natural sister?
c. Can a person marry their first cousin?
d. Can a man marry his widow's sister?
e. Can a woman marry her uncle?

These questions tap into your knowledge about what the marriage laws are. But they do more than that. One of the questions usually gets the answer 'yes', when in fact the true answer has to be 'no': Can a man marry his widow's sister? The answer has to be 'no', because a person who has a widow is dead.

Anomalies such as these show that word meanings are not necessarily fully analysed and/or integrated into the mental representation of a discourse (which of these alternatives holds is an empirical issue). There are many such cases (see Sanford and Sturt, 2002), the best known being the so-called Moses illusion (Erickson and Matteson, 1981):

- Moses put two of each sort of animal on the Ark. True or False? Give your answer before reading on.

The answer, of course, is that it is false because *Noah* was the one with the Ark.

It is perhaps even more interesting that the extent to which meaning is processed depends upon the syntactic construction of sentences. Baker and Wagner (1987) presented participants with sentences like (16) and (17), and asked them to say whether the statements were true or not. Try the first one for yourself:

(16) The liver, which is an organ found only in humans, is often damaged by heavy drinking.

(17) The liver, which is often damaged by heavy drinking, is an organ found only in humans.

Participants spotted that sentences like (16) were false about 69 per cent of the time; here, the statement '*an organ found only in humans*', which is of course false, is in the subordinate position. In (17), where it is in the main clause, errors were spotted 80 per cent of the time. So, putting information in a subordinate clause makes it less detectable than if it is in the main clause. Practically, if you want someone not to scrutinize what you are saying, put the bit that you want them to miss in a subordinate position!

Subordination is an example of how syntax controls the extent of processing. There is also evidence that the part of the sentence on which focus or emphasis is put affects the depth of semantic processing. A **cleft sentence** has a structure like *It was John who opened the door*, with *it was* being one half of the cleft. Using *It was…* clearly indicates that the sentence answers the question *Who opened the door*? Using standard Moses illusion sentences, Bredart and Modolo (1988) showed that in (19), detection of the anomaly was much better than with (18):

(18) Moses put two of each kind of animal on the Ark. (True or False?)

(19) It was Moses who put two of each kind of animal on the Ark. (True or False?)

So, the focused elements of a sentence appear to experience deeper processing than the non-focused elements. These simple observations open the way to developing processing theories in a new direction, by showing how the forms of sentences control the amount of processing effort afforded the retrieval of meaning from words. There are many other situations in which shallow processing may occur, and establishing these will enable us to give more sophisticated accounts of language comprehension. Papers by Ferreira *et al.* (2002) and Sanford and Sturt (2002) illustrate the scale of the problem.

2.2.2 Perspective and spin in communicating quantities

Language provides perspective, and so controls the way we reason about things. When we read a novel, for instance, we typically take the perspective of the principal character. Perspective effects are found everywhere in language, and represent an important phenomenon for theories of understanding. However, some of the very simplest cases also have practical consequences for us all.

There is a growing interest in how to communicate everyday risks more effectively. For instance, medicines may have side-effects, and how they are described influences our perception of the risks involved. Similarly, descriptions of foodstuffs may be slanted to make the foodstuff sound as healthy as possible. Recent work from several investigators poses some interesting challenges, both practical and theoretical.

Look at the different ways that the fat content of a food might be portrayed to a customer:

- Contains 9% fat
- Contains less than 10% fat
- 90% fat-free

If you were trying to sell this product, which description would be best? By far the most common formulation is *% fat-free*. Levin and Gaeth (1988) found that describing minced beef as *75% lean* rather than *25% fat* led people to rate the beef as leaner, less greasy, and of higher quality, an effect that lasted even after they had tasted the beef! In much the same way, the fat-free formulation above draws attention to leanness, while the percentage-fat formulation highlights the fact that there is fat. Sanford *et al.* (2002) studied the basis of this phenomenon. They noted that if people think that fat is unhealthy, the formulations lead to different evaluations. So, intuitively, in the first case below, *bad thing* is acceptable, whereas in the second, *good thing* is (although both depict the same amount of fat):

> *(20) This product contains 10% fat, which is a bad thing.*
>
> *(21)This product is 90% fat-free, which is a good thing.*

In a reading time experiment, participants saw materials like this:

> *A new home-made style yoghurt is to be sold in supermarkets.*
>
> *The yoghurt [contains 5% fat/25% fat] / [is 95% fat-free/is 75% fat free].*
>
> *It is widely believed to be a healthy/unhealthy product.*

The brackets and slashes indicate alternative conditions options. So, for instance, the *5% fat* formulation in the fat statement sentence could be followed by either the *healthy* or the *unhealthy* continuation in the next sentence.

Participants read texts like these, one sentence at a time, using a self-paced procedure. The prediction was that if a description makes a product sound healthy, then the ‘healthy product’ version of the final sentence should be read more quickly than the ‘unhealthy product’ version, because it would make more sense and be easier to integrate. The results are shown in Figure 7.2. When the percentage fat formulation is used, *5% fat* leads to faster reading times for the healthy version. This pattern changes for *25% fat*: reading time for ‘healthy’ goes up, while reading time for ‘unhealthy’ goes down. So, *5% fat* is taken as healthier than *25% fat*, as one might expect. People are bringing their knowledge of expected amounts of fat to bear on the situation. But look at the results for fat-free. Both *95% fat-free* and *75% fat-free* lead to faster integration of the ‘healthy product’ target, and there is no difference between these two. So, given the fat-free formulation, people appear *not* to be using their

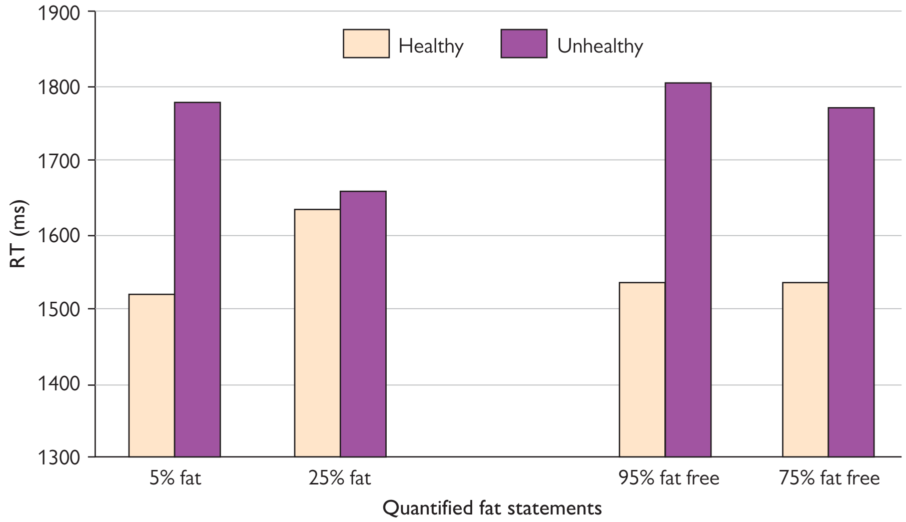


FIGURE 7.2 Effect of quantified fat statement upon acceptance of healthy/unhealthy continuation. *Source: Sanford et al. (2002)*

knowledge. Sanford *et al.* (2002) suggested that the fat-free formulation effectively stops people from utilizing the kind of knowledge they would use with the percentage-fat formulation. Other formulations, such as *less than 10% fat*, have not been investigated at the time of writing.

These are very clear effects of perspective: in the case of percentage fat-free, the perspective is on the amount of non-fat (or healthy) ingredient, whereas in the case of percentage fat, the perspective is on the amount of fat. Although these two formulations actually depict the same amount of fat, they lead to quite different mental operations during understanding.

Perspective effects like this are subtle, but influence how we think of things. With risks, for instance, compare the following formulations:

- Side-effects, including headaches, occur rarely.
- Side-effects, including headaches, occur occasionally.

Although *rarely* and *occasionally* both denote headaches occurring a small proportion of the time, you could fit the continuation *which is a good thing* to the first sentence, and *which is a bad thing* to the second one. They point to different perspectives, so that the same chance of a side-effect can sound *good* or *bad*.

SUMMARY OF SECTION 2

In the first part of the chapter we have seen various examples of how we use language affecting how we interpret the meaning of what is written. Because language is such an important tool for communication it is not surprising that language use has such an impact on our judgements about things, such as how risky they are, as well as our ability to understand and solve basic problems.

In the second part of the chapter we will again look at language in action, but in the context of speech and dialogue. Again, we shall find that language and the circumstances of language usage can play an important role in activities not normally associated with language processing, such as group decision making.

3 LANGUAGE PRODUCTION AND DIALOGUE

Up until now we have considered language processing purely in terms of language comprehension. This reflects the history of psycholinguistics, which for many years treated language processing as equivalent to language comprehension. However, the primary setting for language use is in dialogue and dialogue highlights the importance of language production. Over the last 15 or so years there has been growing interest in production and even more recently in the dynamics of dialogue. The second part of the chapter concentrates on language in action, first in relation to production as a self-contained process and then in relation to both production and comprehension as they occur in dialogue.

3.1 Language production as a self-contained process

For any competent speaker language production seems a straightforward process. For instance, when holding a conversation you are rarely aware of encountering any difficulty in formulating your utterances. However, the apparent ease of language production in informal settings, such as during a conversation, disguises the fact that it is a complex multi-stage process. The complexity is more apparent when writing. For instance, imagine that you have to write a report on this chapter of the book. Suddenly language production becomes difficult. You may have trouble finding the right words to express yourself, or have problems organizing the report into a readily understandable document; you might even have trouble producing strictly grammatical sentences. Here we introduce the language production process in both these situations. First, we consider production as a self-contained process, as when you have to produce something like that report. Then, we consider why it seems to become more straightforward in the informal setting of a dialogue.

3.1.1 Overall architecture of the language production system

Much of what is known about language production has come from the study of speech errors. So first we consider what speech errors can tell us about the overall organization of the language production process, and then we consider in more detail recent work on two particular topics – how speakers design their utterances for particular listeners and how speakers monitor their spoken output.

3.1.2 Speech errors

Speakers make relatively few errors in normal speech (roughly 1 in every 2,000 utterances contains an error), but the errors they do make provide useful evidence about the overall organization of the language production system. Table 7.1 shows the range of different kinds of speech errors that have been regularly observed.

At first sight it looks as if almost any kind of error can occur, but closer examination reveals interesting limitations. Take, for example, the exchange errors. It turns out that exchange errors always occur between items of the same category: nouns exchange with other nouns, verbs with other verbs; and when phonemes (minimal units of speech sound) are exchanged it tends to be consonants with consonants and vowels with vowels. This immediately suggests that the choice of the linguistic units in formulating the utterance (e.g. choice of words or choice of phonemes) is distinct from the grammatical formulation of the utterance in terms of the ordering of those units. In other words, the choice of the word or phoneme occurs at a separate stage from the decision about where the word or phoneme should be placed in the utterance sequence. Otherwise, it would be difficult to explain how the units could end up so far away from where they were supposed to be. Another striking phenomenon is what is called *stranding*. Take example 7 in

TABLE 7.1 Sample speech errors (from Bock and Huitema, 1999)

Type of error	Intended utterance	Error
1. Word anticipation	bury me right with him	bury him right with him
2. Sound anticipation	the lush list	the lust list
3. Word perseveration	evidence brought to bear on representational theories	evidence brought to bear on representational evidence
4. Sound perseveration	President Bush's budget	President Bush's boodget
5. Word exchange	the head of a pin	the pin of a head
6. Sound exchange	occipital activity	accipital octivity
7. Stranding exchange	the dome doesn't have any windows	the window doesn't have any domes
8. Phrase exchange	the death of his son from leukaemia	the death of leukaemia from his son
9. Semantically related word-substitution	I like berries with my cereal	I like berries with my fruit
10. Phonologically related word-substitution	part of a community	part of a committee
11. Sound substitution	the disparity	the disparigy
12. Word blend	it really stood/stuck out	it really stook out
13. Phrase blend	at large/on the loose	at the loose

Table 7.1. The speaker had intended to say *The dome doesn't have any windows*, but the plural 'windows' was placed in the part of the sentence where the singular word 'dome' should have been and vice versa. What came out was not *The* ***windows*** *doesn't have any* ***dome*** but rather *The* ***window*** *doesn't have any* ***domes***. Hence, the assignment of the number feature on the words *window* and *dome* (i.e. whether they were marked as singular or plural) seems to involve a separate stage in the process from the choice of word and placement of that word in the sentence. Another such example of stranding is when the speaker intended to say *If that was done to me* and it came out as *If* ***I*** *was done to that* and not *If* ***me*** *was done to that*. In this case it is the grammatical case marking of the pronoun (i.e. *I* for subject of the verb *was done* and *me* for object) that is assigned separately after the pronoun has been put into that position in the sentence. So the examples indicate that speech errors are more subtle than they seemed at first. Also, they tell us something about what sorts of operations go together in producing an utterance.

On the basis of such patterns of errors the overall organization of the language production system is commonly viewed as in Figure 7.3 (see Bock, 1996 and Levelt, 1989 for a full account). It has three main processes. First, there is the process of formulating the message in a pre-linguistic form. Then, there are two distinct processing systems, the first concerned with formulating the grammatical aspects of the utterance and the second concerned with converting this into the appropriate sequence of sounds (the phonological component). There are many reasons for making this split. First, speech errors are predominately either lexical errors or errors involving phonemes. In fact, the first six kinds of error in Table 7.1, anticipation, preservation, and exchange, can either be lexical (e.g. 1, 3, 5) or phonological (e.g. 2, 4, 6), but rarely involve other units of speech. In other words, even though morphemes (i.e. meaningful bits of words like the *meaning* and the *-ful* in meaningful) are more prevalent than words, there are very few morphemic errors. Similarly, errors involving the more common phonetic features (i.e. sound segments out of which the phonemes are made) are much rarer than errors involving phonemes. This suggests that the two main processes either arrange words (grammatical encoding) or arrange phonemes (phonological encoding). Another piece of evidence relates to the earlier point about lexical exchanges in which nouns exchange with nouns and verbs with verbs. If grammatical and phonological encoding occurred together we would expect phonological exchanges to occur within words of the same grammatical class.

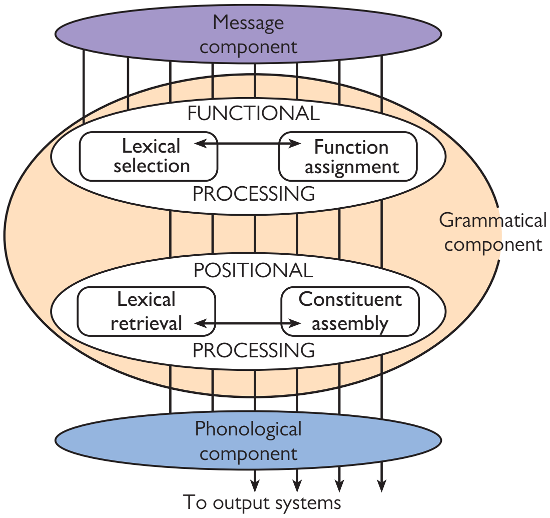


FIGURE 7.3 A summary of the organization of the language production system. *Source: Bock, 1995*

However, it turns out that the grammatical category of the words in which phonological exchanges occur can be completely variable. In other words, grammatical class is only relevant to lexical exchange errors and not to phonological exchange errors. Finally, it turns out that the different kinds of exchange error, lexical and phonological, occur across spans of quite different length. Phonological exchanges tend to be from neighbouring words, whereas lexical exchanges tend to occur across neighbouring phrases (see Box 7.2). This would suggest that the two kinds of process, grammatical and phonological, operate across rather different domains. Whereas grammatical processing can take a long-term view of the utterance, phonological processing only operates locally, one word or two words at a time.

So there is a basic distinction between the grammatical and phonological encoding processes. Looking in more detail at the model we can see that each of these also involve distinct operations. For example, in grammatical processing there are operations that select the word to be uttered (lexical selection) and operations that determine its semantic function in the sentence (e.g. if it is a noun, whether it is to be the agent of the verb, the person that carries out the action, or the patient, the person acted upon). Still within the grammatical component, there is another process that recovers the word form and another that builds up the grammatical constituents of the utterance. Again, evidence for this distinction can be found in the speech errors. Stranding errors, such as example 7 in Table 7.1, indicate that word selection occurs separately from retrieving the precise form of the word. In *The window has several domes*, an abstract representation of the word *window* is selected and put into the sequence before its precise form as *window* is retrieved, hence it can end up as *window* rather than the intended *windows*.

At the next level down in the system, the phonological component, it is also claimed that there are different sub-components to deal with phonemes on the one hand, and larger syllabic units that carry the stress patterns in the speech on the other. We could go into much more detail about the organization of the language production system, but it is beyond the scope of a general cognition course. Instead, let's consider a couple of special topics in more detail.

BOX 7.2 Tongue twister exercise, based on Wilshire (1999)

Try repeating the words below quickly. Start by repeating pod-cab etc. and then after a while have a go with moss-knife etc. If possible record your repetitions on tape and then examine them for different kinds of phonological speech error.

POD CAB CORD PUB . . .

MOSS KNIFE NOOSE MUFF . . .

You should notice that you will start to make errors that involve substituting the speech sounds from one word into the next. Typically, the errors will be anticipations where you might say *nuss* for *noose*, anticipating the following *muff*. This exercise illustrates how phonemes can be substituted in adjacent words during normal speech.

3.1.3 Message selection and audience design

In looking at the language production process we have adopted a similar model to that assumed in the work on language comprehension discussed earlier in Chapter 6. Basically, production when viewed as an isolated process is seen as a kind of translation from an idea via intermediate levels to a sequence of articulated sounds. But of course when we use language to communicate, the speaker has to make a number of more complicated general decisions about how to formulate what he or she wants to say in such a way that it will make sense for that particular listener. This general topic is what has been called audience design.

Audience design is an interesting part of the production process because it requires the speaker to draw complicated inferences about what the listener knows. These inferences are more complicated than you might imagine because they usually involve establishing what is called common ground. Technically common ground relates to the knowledge that the speaker and listener share *and that they both know that they share* (Clark, 1996). Common ground is important because it affects how you should formulate your utterance in such a way that you can be sure it will be understood as you intended it to be. For example, say you are going around an art gallery with a companion and you turn to gaze at a painting that you really like. You might say to them, 'It's great, isn't it?' Now, under certain circumstances that would be a perfectly felicitous statement and convey to your companion that you really liked that particular painting. However, under other circumstances it would be totally uninformative for them. It all depends on what you know that they know at that time, and what you know that they know that you know, and so on. For instance, if you can see from the corner of your eye that they are looking at the same painting and that they can see that you are also looking at that painting, then the statement is quite clear. Both of you take the thing that is really great to be the painting that you both know that you are both looking at. However, on other occasions the statement may not be felicitous. Anything that blocks establishing common ground could lead to misunderstanding. For instance, if you were standing looking at the painting but there was a barrier obscuring your view of your companion, then they would have no basis for establishing what it was you intended to speak about in saying 'It's great, isn't it?' Even if the barrier were a one-way mirror allowing you to see them but not allowing them to see you or what you were looking at, the statement would be infelicitous.

This might seem to be a rather special and complicated example, but audience design enters into almost everything we say. Take for example an experiment by Isaacs and Clark (1987). They used a situation called the referential communication task shown in Box 7.3 (Krauss and Weinheimer, 1967). However, in their case the director (see Box 7.3) had to indicate to the matcher which picture he or she was looking at from a set of pictures of buildings in New York City. In effect, the director had to describe the picture unambiguously to the matcher. Now the twist in the experiment was that the communicating pairs were chosen such that either one or both or neither were New Yorkers. In other words, the common ground between the different pairs could be quite different and the question was whether or not the directors would alter the design of their descriptions according to their assessment of common ground. It turned out that everyone was extremely good at the task and almost immediately established whether they were both New Yorkers, only one was a New Yorker, or neither were. In fact 85 per cent of pairs knew into which category they both fell after only describing two of the cards. But perhaps more interesting was how the speakers then adapted their descriptions according to whom it was that they were speaking to. New Yorkers speaking to New Yorkers would typically just name the building in the picture (e.g. 'Chrysler building') whereas New Yorkers speaking to non-New Yorkers would describe the picture itself (e.g. 'Well, it has three buildings in it, a tall one and two short ones') and so design their utterances according to what they knew was the common ground.

This experiment illustrates nicely how language production in a real context involves much more than just translating ideas into sounds. It also requires a complex assessment of what the listener knows at the time, even including what is in common ground

BOX 7.3 The referential communication task (from Glucksberg and Danks, 1975)

In this task there is a director (speaker) and a matcher (listener) who are separated by a thin partition. The director picks wooden blocks impaled on a stake and has to communicate to the matcher the nonsense pattern shown on the block (see the right side of Figure 7.4 for examples). The matcher then chooses the appropriate block from his or her pile and puts it onto their stake. With this task it is possible both to analyse what the communicators say to each other, and to establish how accurately they can communicate the patterns by comparing the order of items that the director started out with to the order of items that the matcher ended up with. You could try it out with your friends.

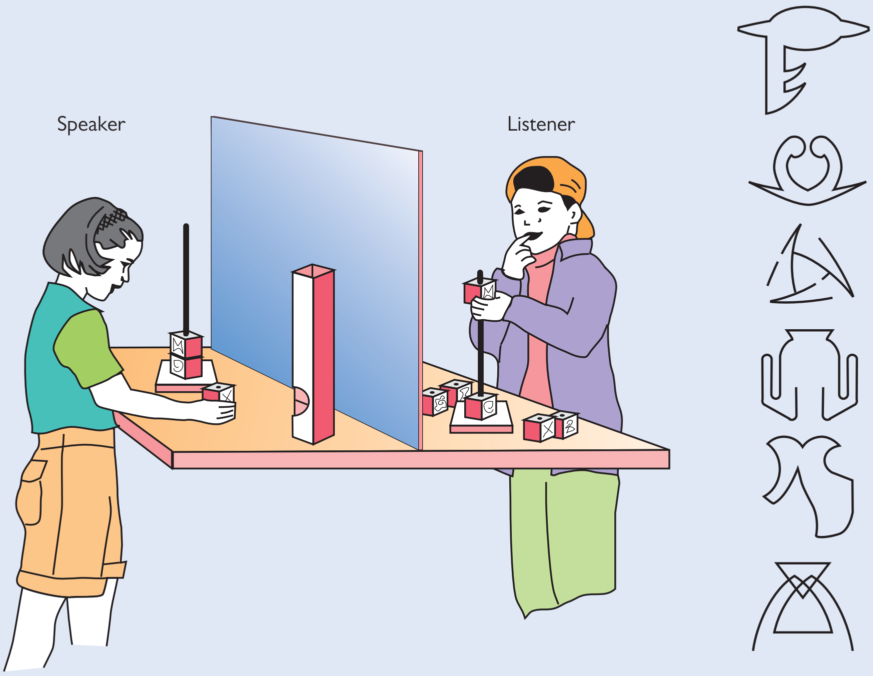


FIGURE 7.4 The referential communication task. *Source: Glucksberg and Danks, 1975*

between the speaker and listener. However, there is controversy over the extent to which speakers always take common ground into account. For example, Horton and Keysar (1996) found that speakers under time pressure did not produce descriptions that took advantage of what they knew about the listener's view of the relevant scene. In other words, the descriptions were formulated with respect to the speaker's current knowledge of the scene rather than with respect to the speaker and listener's common ground. Similarly, Keysar *et al.* (1998) found that in visually searching for a referent for a description, listeners are just as likely to initially look at things that are not part of the common ground as things that are. In other words, listeners also do not seem to always take advantage of common ground.

Nevertheless, Horton and Keysar (1996) found that with less time pressure speakers often did take account of common ground in formulating their utterances, and Keysar *et al.* (1998) argued that listeners at a later monitoring stage take account of common ground in comprehension (see also Brennan and Clark, 1996; Schober and Brennan, 2003). For a more detailed review of this controversy see Garrod and Pickering (2007).

So all in all it seems that audience design, and the extent to which the audience as comprehender is sensitive to design, is a complicated issue. In the absence of time pressure both language producers and language comprehenders are able to take into account their common ground in processing an utterance. However, when under time pressure this kind of complex assessment of listener by speaker and vice versa is one of the first parts of the process to suffer. Below we consider how this separation of processes in production may depend upon the distinction between initial formulation of an utterance and subsequent monitoring and correction of that utterance.

3.1.4 Self-monitoring

An important part of the process of speech production is being able to monitor and correct what you are saying (Hartsuiker and Westerberg, 2000). We know that speakers are always doing this because natural speech is full of minor hesitations and dysfluencies in which the speaker briefly stops and corrects or **repairs** their utterance.

There is a real issue as to how this monitoring process operates. In its most straightforward form monitoring can work by the speaker listening to and comprehending their own output. Then, as soon as they encounter something that does not match what they had originally intended to say, they can stop the speech, reformulate the utterance, and continue with the repaired fragment. For example, take turn 6 in the dialogue extract shown in Section 3.2.1. Here the speaker starts out saying 'The left . . .' but then realizes that he intended to be more explicit. So he stops and restarts his utterance with the repair 'going from left to right in the second box'. It is straightforward to assume the existence of such an 'outer-loop' monitoring process because we know that we can perfectly well understand what we are saying to someone else. More controversial is the idea that monitoring can also operate at earlier stages in the production process. For instance, that it can operate at the level of the message formulator, or later at the stage of phonological encoding.

Evidence for this kind of 'inner-loop' monitoring comes from a variety of sources. The most colourful evidence is from a speech error elicitation experiment by Motley *et al.* (1982). They used a device for producing speech errors that was a bit like the tongue-twister elicitation procedure described in Box 7.2. Participants had to repeat context sequences that were likely to produce errors in subsequent critical items, such as the pair 'barn door', which might be mispronounced as 'darn bore'. However, they also included critical pairs of items that if mispronounced would lead to a taboo word (e.g. 'tool kits' as 'cool ???'). The crucial question was whether speakers were as likely to come up with the taboo forms as they were to come up with non-taboo forms. Motley *et al.* found that the taboo errors were much less likely to occur than non-taboo errors. They argued that the taboo errors must have been filtered out before the utterance had been articulated. Now, it is difficult to understand how such a pre-articulatory filtering could occur without some form of 'inner-loop' monitoring; that is, monitoring based on something available before the speaker has had a chance to listen to what they are saying.

Another kind of evidence comes from examining the temporal characteristics of speech errors and their corrections. The crucial measure is the time between producing the incorrect word and producing the repair (e.g. the time between saying 'the left . . .' and 'going from left . . .' in utterance 6 in the example in Section 3.2.1). Hartsuiker and Kolk (2001) argued that any error correction time interval of less than 150 ms could not reflect the outer-loop monitoring process, because this would not allow sufficient time to comprehend the output, reformulate it, and then restart the utterance. Looking at several collections of such errors they found a high preponderance of these short latency restarts. In fact, their results suggested that the majority of speech repairs were based on 'inner-loop' as opposed to 'outer-loop' monitoring. An additional interesting feature of these two kinds of monitoring is that whereas outer-loop monitoring appears to tax attention, 'inner-loop' monitoring does not. This explains why when people speak under time pressure they tend to produce fewer overt speech repairs than when they speak more slowly, and the overt repairs tend to involve short latency restarts. It is assumed that with high time pressure, monitoring and repair shift even more in favour of the low attention 'inner-loop' route. This is also one reason why Keysar *et al.* (1998) argued that audience design might only influence production at a later stage on the basis of self-monitoring. The idea was that such monitoring would be associated with the outer loop and so be less effective when speaking under time pressure.

A similar distinction between 'inner-loop' and 'outer-loop' monitoring is made for monitoring of physical movements, such as when grasping something or when picking something up (see Blakemore *et al.*, 2002). Again, it seems that the inner loop equivalent for motor control is not accessible to consciousness and does not tax attention to the same degree that the outer loop system does.

3.1.5 Summary

In this section we looked at two special topics in language production that are outside the basic production processes discussed earlier in Sections 3.1.1 and 3.1.2. The first of these special topics, audience design, was about ways in which the speaker attempts to take into account the listener during production. As we saw, there is some controversy over the degree to which audience design, at least in relation to taking into account the common ground, is an obligatory part of the normal language production process. Whereas Clark and colleagues have argued that speakers aid their listeners by taking account of common ground when formulating their utterances, others have argued that common ground is only taken into account on the basis of an optional outer-loop monitoring and repair process.

The second special topic we looked at was the mechanisms of self-monitoring in speech. The important issue here is whether such monitoring can only occur on the basis of a speaker listening to himself or herself – outer-loop monitoring – or whether it also proceeds on the basis of pre-articulatory monitoring – inner-loop monitoring. Evidence from the time course of self-repair strongly indicates the prevalence of pre-articulatory inner-loop monitoring.

3.2 The challenge of dialogue

So far we have only considered language production as if it were a process completely isolated from comprehension. However, language processing most often occurs in the context of a dialogue where each participant both produces and comprehends more or less at the same time. How does this kind of interaction affect the production and comprehension process?

First, we take a look at what sort of thing goes on in dialogue and how the language itself is different in dialogue from monologue. This then leads to a more general discussion about how dialogue and monologue involve different kinds of processing, and in turn how this influences the nature of production and comprehension in a dialogue setting. Finally, we consider a recent model of language processing in dialogue that takes these differences into account.

3.2.1 What is dialogue?

The following example comes from a transcript of two players in a cooperative maze game. One player, *A*, is trying to describe his position to his partner, *B*, who is viewing the same maze on a computer screen in another room (see Garrod and Anderson, 1987). The position being described in all the utterances shown in bold is illustrated schematically in Figure 7.5.

1——B: Tell me where you are?
2——A: Ehm . . . oh God. *(laughs)*
3——B: *(laughs)*
4——A: Right: **two along from the bottom, one up.**
5——B: Two along from the bottom, which side?
6——A: The left: going from left to right in the second box.
7——B: You're in the second box.
8——A: One up. *(1 sec.)* I take it we've got identical mazes?
9——B: Yeah, well: right, starting from the left, **you're one along.**
10——A: Uh-huh.
11——**B: and one up?**
12——A: Yeah, and I'm trying to get to . . . *etc.*

At first glance the language looks disorganized. Strictly speaking many of the utterances are not grammatical sentences – only one of the first six contains a verb. There are occasions when production of the same sentence is shared between the speakers, as in utterances (7–8).

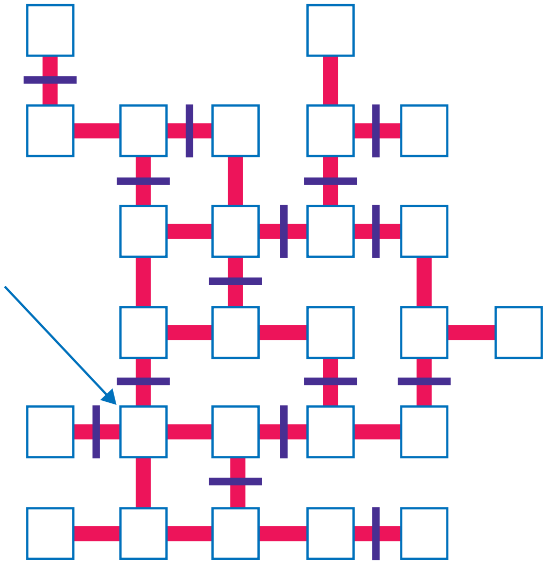


FIGURE 7.5 The arrow points to the position on the maze that *A* and *B* are trying to describe in the dialogue extract shown in the text. Notice that the two descriptions in the text are in fact different – *Two along from the bottom one up* vs *One along . . . one up.*

In fact, the sequence is quite orderly so long as we assume that dialogue is a joint activity (Clark, 1996). In other words, dialogue involves cooperation between interlocutors in a way that allows them to sufficiently understand the meaning of the dialogue as a whole, and this meaning results from these joint processes. So dialogue is orderly once it is recognized that it requires coordination to establish consensus between the two speakers.

3.2.2 Dialogue and consensus

In a piece of written text, whether it is a newspaper article or the chapter of a learned volume, the meaning is there on the page waiting to be extracted. If it is well written and you are a competent reader, then you should be able to come to an interpretation that matches roughly what the writer intended (see Section 2). However, this does not depend on establishing any kind of consensus with the author. After all, he or she may well be long dead and gone.

In dialogue the situation is very different. Dialogue is organized around establishing consensus. First, dialogue turns are linked across interlocutors (Schegloff and Sacks, 1973). An imperative question, such as (1) 'Tell me where you are?' calls for a response, such as (4) 'Two along from the bottom, one up'. Even a statement like (4) 'Right: two along from the bottom, one up' cannot stand alone. It requires either an affirmation or some form of query, such as (5) 'Two along from the bottom, which side?' This means that production and comprehension processes become coupled. *B* produces an imperative question and expects an answer of a particular type; *A* hears the question and has to produce an answer of that type. For example, after saying 'Tell me where you are?' in (1), *B* has to understand 'two along from the bottom, one up' in (4) as a reference to *A*'s position on the maze; any other interpretation is ruled out.

Second, the meaning of what is being communicated depends on the interlocutors' agreement or consensus rather than on dictionary meanings and is therefore subject to negotiation. Take for example utterances (4–11) in the dialogue example above. In utterance (4), *A* describes his position as 'Two along from the bottom, one up' but the final interpretation is only established at the end of the first exchange when consensus is reached on a rather different description by *B* (9–11): 'You're one along . . . and one up?'

So in dialogue the interpretation depends upon taking part in the interaction itself. This was nicely demonstrated in an experiment by Schober and Clark (1989). They used an experimental set-up similar to that used by Krauss and Weinheimer (1967) (see Box 7.3), in which a director had to describe a sequence of abstract Chinese tangram patterns to a matcher on the other side of a screen. However, in this experiment there was also a third person who overheard everything that was said but could not interact with the director. The overhearer had to try and pick the cards being described by the director in the same way that the matcher had to pick them. It turned out that overhearers, who could not interact with the director, performed consistently less well than the matchers who could interact with the directors. Schober and Clark argued that overhearers are at a disadvantage because they cannot control what the director is saying in the way that the participant can (e.g. a participant can always query what they fail to understand whereas an overhearer cannot). So hearing everything that is being said does not lead to a full understanding when you cannot interact directly with the speaker.

The third way that dialogue involves coordinated processing relates to the general problem of ambiguity discussed in Chapter 6. In the extract shown above, the participants spend all the time trying to work out a mutually acceptable and unambiguous description for *A*'s location on the maze. As we shall see below, this is achieved through a process of coordinating outputs with inputs: speakers always attempt to generate utterances that correspond semantically to the utterances that they have recently had to comprehend. Consider for example utterances 4 and 5 above, in which *B* echoes the description 'Two along from the bottom . . .' As a result of such output-input coordination, the same expression comes to take on the same precise meaning within any stretch of dialogue.

Finally, dialogue participants try to establish a coordinated conception of their topic. In the case of the maze game illustrated above, this amounts to converging on a common spatial concept of the maze's configuration. Thus some people playing this game will refer to their locations by reference to *right-turn indicators*, *upside-down T shapes*, or *Ls on their sides*. These speakers, unlike the pair responsible for the dialogue illustrated above, conceive of the maze as a conglomeration of patterns or shapes, each with a different name. Conversational partners often establish quite idiosyncratic conceptions of the topic (as with the use of *right-turn indicator* or *upside-down T shapes*) but in well-managed dialogues they always align on the *same* idiosyncratic conception. Again, this process supports consensus, which is the fundamental goal of dialogue.

One of the reasons why dialogue presents such a challenge to processing accounts is that these interactive characteristics are difficult to reconcile with the standard view of communication as a one-way process of *information transfer*. And it is just such a view that underpins much of the work in psycholinguistics (such as the work described in Chapter 6). Here we argue that a more useful processing framework for dialogue may be based on the notion of *interactive alignment*. According to this account, dialogue participants come to align their linguistic representations at many levels. The alignment process helps them to come to a mutually satisfactory interpretation of what is being said and it greatly simplifies the basic processes of production and comprehension during dialogue.

3.2.3 A model of dialogue processing

The interactive alignment account starts with the simple observation that dialogue participants alternate between speaking and comprehending. Furthermore, the representations that are used for comprehension (whether they be syntactic, lexical, or at the level of articulation) will activate or prime matching representations in production.

If we assume that representations active during comprehension remain active during subsequent production, then there will always be a tendency for interlocutors to coordinate outputs (productions) with inputs (what has just been understood). Put two such systems together in a dialogue and the overall system will only be completely stable if the two adopt aligned linguistic representations at every level (see Garrod and Pickering, 2004). Pickering and Garrod (2004) go on to argue that this kind of interactive alignment of representations supports mutual understanding because alignment does not just occur within independent levels of the system, but also serves to link those levels with each other. In other words, the automatic alignment of representations at all levels will tend to establish a kind of common ground between the two communicators, which aids mutual understanding.

First we look at the evidence for representational alignment during dialogue and consider how it may support the interpretation process. Then we go into a little more detail on one consequence of representational alignment in relation to language processing, which is what has been called *routinization*.

3.2.4 Evidence for interactive alignment

Dialogue transcripts are full of repeated linguistic elements and structures indicating alignment at various levels (Aijmer, 1996). Alignment of lexical processing during dialogue was specifically demonstrated by Garrod and Anderson (1987), as in the example dialogue in Section 3.2.1, and by Clark and colleagues (Brennan and Clark, 1996; Wilkes-Gibbs and Clark, 1992). These latter studies show that interlocutors tend to develop the same set of expressions to refer to particular objects, and that the expressions become shorter and more similar on repetition with the same interlocutor but are modified if the interlocutor changes.

Levelt and Kelter (1982) found that speakers tended to reply to 'What time do you close?' or 'At what time do you close?' (in Dutch) with a congruent answer (e.g. 'Five o'clock' or 'At five o'clock'). This alignment may be syntactic (repetition of phrasal categories) or lexical (repetition of *at*). Branigan *et al.* (2000) found clear evidence for syntactic alignment in dialogue. Participants took it in turns to describe pictures to each other (and to find the appropriate picture in an array). One speaker was actually a confederate of the experimenter and produced scripted responses, such as 'the cowboy offering the banana to the robber' or 'the cowboy offering the robber the banana'. The syntactic structure of the confederate's description influenced the syntactic structure of the experimental subject's description and it did so much more strongly than in a comparable non-dialogue situation. For instance, if the participant had just heard 'the cowboy offering the robber the banana', they would describe the next picture as 'the clown handing the dancer the balloon', whereas if they had just heard 'the cowboy offering the banana to the robber', they would describe it as 'the clown handing the balloon to the dancer'.

Alignment also occurs at the level of articulation. It has long been known that as speakers repeat expressions, articulation becomes increasingly reduced (i.e. the expressions are shortened and become more difficult to recognize when heard in isolation; Fowler and Housum, 1987). However, Bard *et al.* (2000) found that reduction was just as extreme when the repetition was by a different speaker in the dialogue as it was when the repetition was by the original speaker. In other words, whatever is happening to the speaker's articulatory representations is also happening to their interlocutor's. So the two representations are becoming aligned. There is also evidence that interlocutors align accent and speech rate (Giles *et al.*, 1992; Giles and Powesland, 1975).

Taken together these findings indicate that something rather special happens when we process language in a dialogue setting. The representations called upon in production are already in some sense available to the speaker from his or her comprehension of the prior dialogue. Apart from helping the interlocutors to come to a truly aligned interpretation of what the dialogue is about, it also simplifies the production and comprehension processes themselves. One of the ways in which this may happen is through what has been called **routinization**.

3.2.5 Routinization in dialogue processing

The process of alignment means that interlocutors draw upon representations that have been developed during the dialogue. Thus it is not always necessary to construct representations that are used in production or comprehension from scratch. One particularly important implication is that interlocutors develop and use routines (set expressions) during a particular interaction.

A *routine* is an expression that is 'fixed' to a relatively great extent. First, the expression has a much higher frequency than the frequency of its component words would lead us to expect. Second, it has a particular analysis at each level of linguistic representation. Thus, it has a particular meaning, a particular syntactic analysis, a particular pragmatic use, and often particular phonological characteristics (e.g. a fixed intonation). Extreme examples of routines include repetitive conversational patterns such as *How do you do?* and *Thank you very much*. Routines are highly frequent in dialogue. It has been estimated that up to 70 per cent of words in a standard dialogue occur as part of recurrent word combinations. However, different expressions can be routines to different degrees, so actual estimates of their frequency are somewhat arbitrary. Some routines are idioms, but not all (e.g. *I love you* is a routine with a literal interpretation in the best relationships).

Most discussion of routines focuses on phrases whose status as a routine is pretty stable. However, Pickering and Garrod (2005) also claim that routines are set up 'on the fly' during dialogue as a result of the interactive alignment process. They call this *routinization* and it represents one of the rather special features of language processing in a dialogue as opposed to a monologue setting. For more details on the routinization process during dialogue see Pickering and Garrod (2005).

3.3 The monologue dialogue distinction and group decision making

The first part of this chapter concentrated exclusively on comprehension of written language or monologue comprehension. The second half has concentrated on language production and production and comprehension during dialogue. In describing language production we first treated it as an isolated process. In other words, we only considered production in terms of the production of monologue. In the later sections of the chapter we have seen that language processing in dialogue may be rather different from language processing in monologue. First, it depends upon coordinated processes of production and comprehension, as in answering a question. Second, it seems to involve direct participation from both interlocutors in creating a common understanding of the message. Hence, overhearers cannot fully understand what is being said in dialogue. Finally, both production and comprehension may be governed by an interactive alignment process that leads to routinization.

We opened the chapter by contrasting language use in the contexts of monologue and dialogue. But what are the consequences of this distinction beyond language processing itself? One interesting consequence relates to group decision making. Communication is crucial to group decision making, whether it involves a family deciding to move to a larger house, a parliament deciding on new legislation, or a jury coming to a verdict. It is crucial because there is no other way in which a group can come to a consensus, and group decision, at least in its purest form, depends upon consensus. Remember that one of the main distinctions between dialogue and monologue is the way in which dialogue promotes consensus whereas monologue does not. So it is an interesting question as to what kind of communication process operates within a group and how that might affect the way in which people are influenced by other members of the group in coming to a decision.

Imagine that you are a member of a committee discussing some particular issue at your work. Sometimes you are aware of being highly engaged in discussion with just one or two other members of the committee; it is like a two-party dialogue. At other times you just sit back and listen to what the most vociferous member of that committee is saying. Now, afterwards you happen to bump into someone who had been there and to your surprise you discover that what you thought was the crucial decision is not quite the same as what they thought was crucial. Typically people's views about such things vary quite a lot. The question is what affects those views and how does that relate to the communication process during the meeting itself. Are you going to agree more with the people that you had the interactive discussion with at the meeting, or are you going to be influenced most by the vociferous and dominant member of the group?

The monologue and dialogue models of communication bear differently on this question. According to the information transfer or monologue account, a group discussion can be thought of as a process in which there are a series of monologues in which the current speaker broadcasts information to the rest of the group. Hence, you should tend to be influenced most by the person who says the most. Who you speak to next in the discussion should have no special influence on your views. The interactive alignment or dialogue account makes a very different prediction. The people who should have most influence on your views are those with whom you directly interacted, and there should be no particular reason why you should be influenced by the dominant speaker.

Fay *et al.* (2000) report a study that shows that both of these views are correct, but which applies depends on the size of the group holding the discussion. To test this they had two sizes of groups of students imagine that they were a university disciplinary committee who had to sit down and decide as a group what to do about a complex case of student plagiarism. First, each member of the group read a one-page description of the case, and then before discussing it they each had to rank 14 relevant issues in terms of how important they felt they were to this case. The issues ranged from clearly relevant ones, such as the severity of the plagiarism, to more ambiguous issues, such as the university's responsibility to the student. The groups then discussed the case for about 20 minutes and after the discussion each person again ranked the issues, but now in terms of how important they thought they had been to the group as a whole. By comparing the agreement in the ranking scores between each member of the group after the meeting (after accounting for their pre-meeting agreements), it is possible to determine who has the strongest influence on whom with respect to the 14 key issues. Fay *et al.* then used the transcriptions of the discussions to establish (1) which members of the groups had served as high interaction partners for each group member and which had served as low interaction partners, and (2) which members of each group had been dominant speakers as opposed to non-dominant speakers. For (1) they defined a high interaction partner as a person who was most likely to speak either immediately before or immediately after each member of the group. They were then able to examine the degree to which everyone in the groups had been influenced either by their high interaction partners or by the dominant speakers in the discussion. The results were clear and quite striking. It turned out that in small group meetings with five members, people were only especially influenced by their high interaction partners. There was no additional influence coming from dominant speakers. This is exactly what is predicted by the alignment account, because (1) interactants automatically align with each other, and (2) an overhearer or side participant is not going to be influenced by others' interactions even when they involve the dominant speaker. However, in larger groups of ten exactly the opposite pattern emerged. People in the larger groups were all influenced by the dominant speaker and there was absolutely no effect of interaction (see Figure 7.6). This was as predicted by the monologue or information broadcast account. Additional analysis of the meeting transcripts also supported the idea that in the small groups the utterances and turn pattern was just like that in two-party conversation. The utterances were shorter, there were many more interruptions, and the pattern of speaker turns tended to conform to an *ABABA* pattern with the same two speakers taking alternate turns for extended periods of discussion.

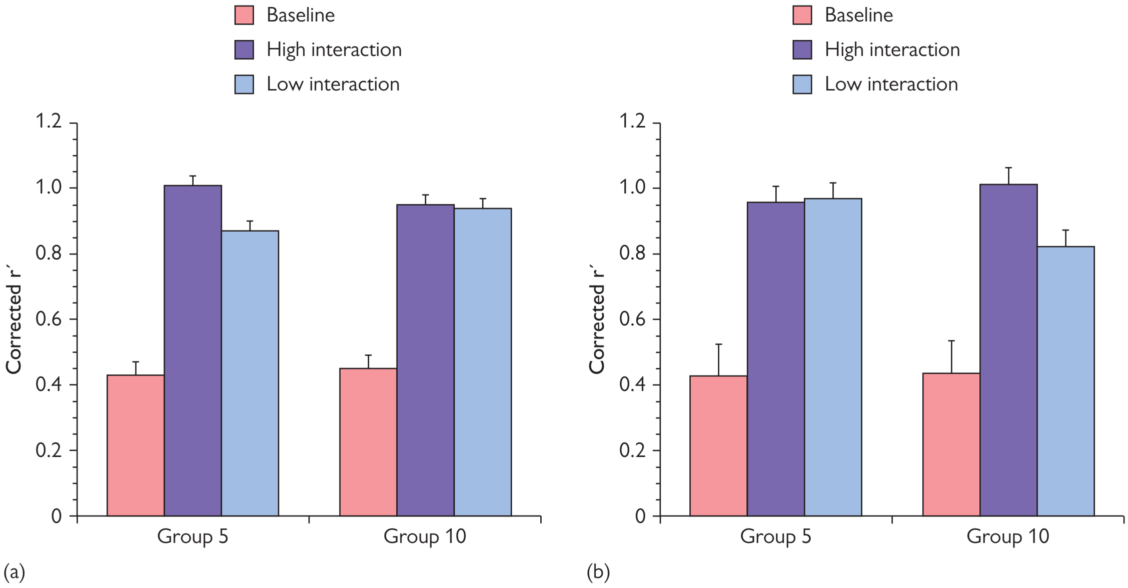


FIGURE 7.6 Correlation between ranking of issues before and after discussion for groups of five and ten discussants. (a) shows correlation between discussants and high and low interaction participants, and (b) shows correlation between the group and dominant speakers and between the group and non-dominant speakers. *Source: Fay et al., 2000*

4 CONCLUSIONS

We began this chapter by arguing that the simple translation account of language processing, whether in terms of comprehension or production, could not capture everything that happens when we use our language to communicate. Language in action involves access to general knowledge, inference beyond what is actually said, and, in the case of dialogue, coordinated action. Without these additional processes communication fails. In fact, it is generally thought that much of the individual differences in reading ability that you encounter in the school classroom or university lecture hall come from differences in readers' abilities to access the appropriate knowledge, and hence their ability to integrate information in the texts they are reading (see Garrod and Daneman, 2003).

In Section 2 we considered these additional processes in relation to text comprehension. We concentrated on the three basic issues that figure in most accounts of comprehension beyond the word: anaphora resolution, non-literal meaning, and text inference. We then went on to discuss in more detail some of the hot topics in the area. So we looked at the argument for embodied representations of meaning, the problem of partial or incomplete processing of text, and the way in which perspective affects interpretation. In all these cases, language comprehension involved more than just the translation of sounds or written symbols into meanings.

In Section 3 we turned our attention to language production and dialogue. Language production when viewed as an isolated process seems to involve the same sort of translation processes assumed for comprehension in Chapter 6. However, when production and comprehension are considered in the context of dialogue the processes take on a different character. The important thing in dialogue processing is how production and comprehension processes become coupled to each other to produce aligned linguistic representations at every level. In turn the contrast between language processing in monologue and dialogue has interesting consequences for such apparently non-cognitive activities as group interaction and decision making.

Having covered some of the ways in which research on language in action has developed over the last 15 years, it is natural to ask where this research might be headed. Of the many possible future directions, we concentrate here on one, which concerns developing a more unified approach to considering the production and comprehension of language. As we pointed out in both Sections 2 and 3, psychologists traditionally study the processes of language production and comprehension in the context of isolated acts of production and comprehension. Yet, as we have argued toward the end of Section 3, language in action typically requires the interweaving of the two. The work on dialogue discussed here also suggests that processes of production and comprehension might be strongly interwoven. In other words, acts of production may call upon comprehension processes and acts of comprehension may call upon production processes. To a certain extent this is already accepted for acts of production, because it is generally assumed that we use the comprehension system to monitor speech production both during outer- and inner-loop monitoring. But what about the use of production processes during acts of comprehension? A recent proposal is that speech comprehension in naturalistic situations calls upon the production system to support predictions about upcoming material (Pickering and Garrod, 2007). This proposal is based first on growing evidence that prediction can play an important role in sentence and discourse comprehension, with predictions operating at all levels including phonemes, words, grammar, and meaning. Secondly, there is increasing evidence for the involvement of production processes (e.g. sub-threshold activation of speech motor programs) during speech comprehension.

A recent neuroimaging study produced intriguing evidence for such interweaving in relation to predictions during naturalistic story understanding. Stevens *et al.* (2010) recorded BOLD activity (i.e. changes in blood oxygenation across the brain) as a participant told an impromptu story in an fMRI scanner, and they recorded the story. Subsequently, they played back the recording to another group of participants while they were also being scanned. Stevens *et al.* then looked for regions in the brain in which the narrator's brain activity correlated with that of the listeners. The results were very striking. First, they discovered that the speaker's brain activity in a wide range of brain areas strongly correlated over time with that of the listeners, and this appeared both in traditional language processing areas and in areas not normally associated with language processing (e.g. medio-frontal cortex). Second, they shifted the time series of brain activity between speaker and listener so that either the speaker's BOLD activity in each region preceded that of the listener or vice versa. This meant that they could determine the extent to which speaker brain activity predicted *subsequent* listener brain activity in any region and the extent to which listener brain activity predicted *subsequent* speaker brain activity in any region. Interestingly, they found evidence for both. Not surprisingly, activity in the speaker's brain often predicted subsequent activity in the listeners' brains, because speakers need to prepare what they are saying and produce it before the listener hears it. More interestingly, they also found the opposite pattern in which activity in the listeners' brains predicted subsequent activity in the speaker's brain. In other words, listeners' brains were often anticipating what was about to happen in a speaker's brain. Of particular interest was the further finding that the degree to which a listener's brain activity predicted the speaker's brain activity correlated highly with the quality of that listener's understanding of the story as measured by analysis of their subsequent recall of the story.

This kind of research linking neural activity in speakers and listeners over time is very much in its infancy. However, it points to exciting new developments in the integrated study of production and comprehension processes when language is truly in action.

FURTHER READING

Fischer, M.H. and Zwaan, R.A. (2008) 'Embodied language: a review of the role of the motor system in language comprehension', *The Quarterly Journal of Experimental Psychology*, vol.61, pp.825–50. This paper gives an up-to-date review of research on embodiment of meaning, which builds on the discussion in Section 2.

Clark, H.H. (1996) *Using Language*, Cambridge, Cambridge University Press. This book gives a much more extensive account of the special characteristics of language use in dialogue. It discusses such issues as dialogue as an instance of joint action and the nature of common ground.

Pickering, M.J. and Garrod, S. (2004) 'Toward a mechanistic psychology of dialogue', *Brain and Behavioral Sciences*, vol.27, pp.169–226. This extended article gives a much more detailed account of interactive alignment in dialogue. The edition also includes a range of commentaries on the article (and responses), which you may find helpful in understanding and evaluating the theory.

REFERENCES

Aijmer, K. (1996) *Conversational Routines in English: Convention and Creativity*, London and New York, Longman.

Baker, L. and Wagner, J.L. (1987) 'Evaluating information for truthfulness: the effects of logical subordination', *Memory & Cognition*, vol.15, pp.247–55.

Bard, E.G., Anderson, A.H., Sotillo, C., Aylett, M., Doherty-Sneddon, G. and Newlands, A. (2000) 'Controlling the intelligibility of referring expressions in dialogue', *Journal of Memory and Language*, vol.42, pp.1–22.

Barton, S. and Sanford, A.J. (1993) 'A case-study of pragmatic anomaly-detection: relevance-driven cohesion patterns', *Memory and Cognition*, vol.21, pp.477–87.

Blakemore, S-J., Wolpert, D.M., and Frith, C.D. (2002) 'Abnormalities in the awareness of action', *Trends in Cognitive Science*, vol.6, pp.237–42.

Bock, K. (1995) 'Sentence production: from mind to mouth', in Miller, J.L. and Eimas, P.D. (eds.) *Handbook of Perception and Cognition: Speech, Language, and Communication*, Orlando, FL, Academic Press.

Bock, K. (1996) 'Language production: methods and methodologies', *Psychonomic Bulletin & Review*, vol.3, no.4, pp.395–421.

Bock, K. and Huitema, J. (1999) 'Language production', in Garrod, S. and Pickering, M. (eds.) *Language Processing*, Hove, Psychology Press.

Branigan, H.P., Pickering, M.J., and Cleland, A.A. (2000) 'Syntactic coordination in dialogue', *Cognition*, vol.75, no.2, B13–B25.

Bredart, S. and Modolo, K. (1988) 'Moses strikes again: focalization effects on a semantic illusion', *Acta Psychologica*, vol.67, pp.135–44.

Brennan, S.E. and Clark, H.H. (1996) 'Conceptual pacts and lexical choice in conversation', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.22, pp.1482–93.

Brown, G. and Yule, G. (1983) *Discourse Analysis*, Cambridge, Cambridge University Press.

Clark, H.H. (1996) *Using Language*, Cambridge, Cambridge University Press.

Erickson, T.A. and Matteson, M.E. (1981) 'From words to meaning: a semantic illusion', *Journal of Verbal Learning and Verbal Behavior*, vol.20, pp.540–52.

Fay, N., Garrod, S., and Carletta, J. (2000) 'Group discussion as interactive dialogue or as serial monologue: the influence of group size', *Psychological Science*, vol.11, no.6, pp.481–6.

Ferreira, F., Ferraro, V., and Bailey, K.G.D. (2002) 'Good-enough representations in language comprehension', *Current Directions in Psychological Science*, vol.11, pp.11–15.

Fodor, J. (2000) *The Mind Doesn't Work That Way*, Cambridge, MA, MIT Press.

Fowler, C. and Housum, J. (1987) 'Talker's signalling of "new" and "old" words in speech and listener's perception and use of the distinction', *Journal of Memory and Language*, vol.26, pp.489–504.

Garnham, A. (2001) *Mental Models and the Interpretation of Anaphora*, Hove, Psychology Press.

Garrod, S. and Anderson, A. (1987) 'Saying what you mean in dialogue: a study in conceptual and semantic co-ordination', *Cognition*, vol.27, pp.181–218.

Garrod, S. and Daneman, M. (2003) 'Reading, the psychology of', in Nadel, L. (ed.) *Encyclopedia of Cognitive Science*, London, Nature Publishing Group, Macmillan.

Garrod, S. and Pickering, M.J. (2004) 'Why is conversation so easy?' *Trends in Cognitive Sciences*, vol.8, pp.8–11.

Garrod, S. and Pickering, M.J. (2007) 'Alignment in dialogue', in Gaskell, G. (ed.) *Oxford Handbook of Psycholinguistics*, Oxford, OUP.

Gibbs, R. (1983) 'Do people always process the literal meaning of indirect requests?' *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.9, pp.524–33.

Giles, H. and Powesland, P.F. (1975) *Speech Styles and Social Evaluation*, New York, Academic Press.

Giles, H., Coupland, N., and Coupland, J. (1992) 'Accommodation theory: communication, context and consequences', in Giles, H., Coupland, J., and Coupland, N. (eds.) *Contexts of Accommodation*, Cambridge, Cambridge University Press.

Glenberg, A.M. and Kaschak, M.P. (2002) 'Grounding language in action', *Psychonomic Bulletin and Review*, vol.9, pp.558–65.

Glucksberg, S. and Danks, J.H. (1975) *Experimental Psycholinguistics: An Introduction*, Hillsdale, NJ, LEA.

Glucksberg, S. and Keysar, B. (1990) 'Understanding metaphorical comparisons: beyond similarity', *Psychological Review*, vol.97, no.1, pp.3–18.

Glucksberg, S., Gildea, P., and Bookin, H. (1982) 'On understanding nonliteral speech: can people ignore metaphors?' *Journal of Verbal Learning and Verbal Behavior*, vol.21, pp.85–98.

Halliday, M.A.K. and Hasan, R. (1976) *Cohesion in English*, London, Longman.

Harnad, S. (1990) 'The symbol grounding problem', *Physica D*, vol.42, pp.335–46.

Hartsuiker, R.J. and Kolk, H.H.J. (2001) 'Error monitoring in speech production: a computational test of perceptual loop theory', *Cognitive Psychology*, vol.42, pp.113–57.

Hartsuiker, R.J. and Westerberg, C. (2000) 'Persistence of word order in written and spoken sentence production', *Cognition*, vol.75, B72–B39.

Haviland, S.E. and Clark, H.H. (1974) 'What's new? Acquiring new information as a process in comprehension', *Journal of Verbal Learning and Verbal Behavior*, vol.13, pp.512–21.

Horton, W.S. and Keysar, B. (1996) 'When do speakers take into account common ground?' *Cognition*, vol.59, pp.91–117.

Isaacs, E.A. and Clark, H.H. (1987) 'References in conversation between experts and novices', *Journal of Experimental Psychology: General*, vol.116, pp.26–37.

Keysar, B., Barr, D.J., Balin, J.A. and Paek, T.S. (1998) 'Definite reference and mutual knowledge: process models of common ground in comprehension', *Journal of Memory and Language*, vol.39, pp.1–20.

Kintsch, W. (1988) 'The role of knowledge in discourse comprehension: a construction-integration model', *Psychological Review*, vol.65, pp.163–82.

Krauss, R.M. and Weinheimer, S. (1967) 'Concurrent feedback, confirmation and the encoding of referents in verbal communications; effects of referent similarity and communication mode on verbal encoding', *Journal of Personality and Social Psychology*, vol.4, pp.343–6.

Lakoff, G. (1987) *Women, Fire and Dangerous Things*, Chicago, University of Chicago Press.

Levelt, W.J.M. (1989) *Speaking: From Intention to Articulation*, Cambridge, MIT Press.

Levelt, W.J.M. and Kelter, S. (1982) 'Surface form and memory in question answering', *Cognitive Psychology*, vol.14, pp.78–106.

Levin, I.P. and Gaeth, G.J. (1988) 'How consumers are affected by the framing of attribute information before and after consuming the product', *Journal of Consumer Research*, vol.15, pp.374–8.

McKoon, G. and Ratcliff, R. (1992) 'Inference during reading', *Psychological Review*, vol.99, pp.440–66.

Motley, M.T., Camden, C.T., and Baars, B.J. (1982) 'Covert formulation and editing of anomalies in speech production: evidence from experimentally elicited slips of the tongue', *Journal of Verbal Learning and Memory*, vol.21, pp.578–94.

Pickering, M.J. and Garrod, S. (2004) 'Toward a mechanistic psychology of dialogue', *Behavioral and Brain Sciences*, vol.27, pp.169–90.

Pickering, M.J. and Garrod, S. (2005) 'Establishing and using routines during dialogue: implications for psychology and linguistics', in Cutler, A. (ed.) *Twenty-first Century Psycholinguistics: Four Cornerstones*, Hillsdale, NJ, Erlbaum.

Pickering, M.J. and Garrod, S. (2007) 'Do people use language production to make predictions during comprehension?' *Trends in Cognitive Sciences*, vol.11, no.3, pp.105–110.

Pinker, S. (1994) *The Language Instinct*, New York, HarperCollins.

Sanford, A.J. and Garrod, S.C. (1981) *Understanding Written Language*, Chichester, John Wiley & Sons.

Sanford, A.J. and Garrod, S. (1998) 'The role of scenario mapping in text comprehension', *Discourse Processes*, vol.26, pp.159–90.

Sanford, A.J. and Sturt, P. (2002) 'Depth of processing in language comprehension: not noticing the evidence', *Trends in Cognitive Sciences*, vol.6, no.9, pp.382–6.

Sanford, A.J., Fay, N., Stewart, A.J., and Moxey, L.M. (2002) 'Perspective in statements of quantity, with implications for consumer psychology', *Psychological Science*, vol.13, pp.130–4.

Schegloff, E.A. and Sacks, H. (1973) 'Opening up closings', *Semiotica*, vol.8, pp.289–327.

Schober, M.F. and Brennan, S.E. (2003) 'Processes of interactive spoken discourse: the role of the partner', in Graesser, A.C., Gernsbacher, M.A., and Goldman, S.R. (eds.) *Handbook of Discourse Processes*, Hillsdale, NJ, Erlbaum.

Schober, M.F. and Clark, H.H. (1989) 'Understanding by addressees and over-hearers', *Cognitive Psychology*, vol.21, pp.211–32.

Searle, J.R. (1980) 'Minds, brains and programs', *Behavioral and Brain Sciences*, vol.3, pp.417–24.

Stevens, G.J., Silbert, L.J., and Hasson, U. (2010, in press) 'Speaker-listener neural coupling underlies successful communication', *Proceedings of the National Academy of Sciences.*

Traxler, M.J. and Pickering, M.J. (1996) 'Plausibility and the processing of unbounded dependencies: an eye-tracking study', *Journal of Memory and Language*, vol.35, pp.454–75.

Van Oostendorp, H. and De Mul, S. (1990) 'Moses beats Adam: a semantic relatedness effect on a semantic illusion', *Acta Psychologica*, vol.74, pp.35–46.

Wilkes-Gibbs, D. and Clark, H.H. (1992) 'Coordinating beliefs in conversation', *Journal of Memory and Language*, vol.31, pp.183–94.

Wilshire, C.E. (1999) 'The "tongue twister" paradigm as a technique for studying phonological encoding', *Language and Speech*, vol.42, no.1, pp.57–82.

@ Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

This page intentionally left blank

PART 3

MEMORY

Introduction

8 Long-term memory: encoding to retrieval

Andrew Rutherford, Gerasimos Markopoulos, Davide Bruno, and Mirjam Brady-Van den Bos

9 Working memory

Graham J. Hitch

INTRODUCTION

In Part 3 you will find two chapters dedicated to the topic of memory. Of course, you will have noticed that all the previous chapters have already included explicit references to and implicit assumptions regarding memory processes and/or memory stores. In the chapters of Part 1, the activation and utilization of stored knowledge was frequently invoked in trying to comprehend the processes of attention, perception, and recognition. Similarly in Part 2, stored information (e.g. the mental lexicon) was seen to be essential to understanding categorization, language, and the construction of successful discourse. The fact is that memory of one sort or another is integral to every form of cognition. However, the chapters in Part 3, and also Chapter 17 in Part 6, differ from the other chapters in that they take memory as their focus of interest rather than as an important incidental to some other major topic.

In Chapter 8, 'Long-term memory: encoding to retrieval', the concern is to understand how information gets into and is recovered from memory. More than that, the emphasis is on understanding how different types of encoding and retrieval operations determine what gets remembered and in what form. The quality of memory, it turns out, results from interactions between encoding processes, the kinds of cognitive representations that are constructed, and the types of retrieval operations that act upon those representations in fulfilling whatever goals a person is intent upon. One theme of the chapter is the sheer difficulty of knowing how best to conceptualize memory. A major distinction is seen between the noun *memory* and the verbs *memorizing* and *remembering* or *recollecting*. That is, on the one hand, memory can be conceived as a set of stores and, on the other, memory can be thought of as a set of systems or processes. As you will see there are arguments and data that favour and count against both conceptualizations. Whichever conceptualization one opts for, there is then the problem of deciding how many stores or how many processes to postulate.

One reason that these questions can be so hard to answer is introduced at the start of Chapter 8. It is that the functions of memory in normal everyday cognition are so vast and diverse, and for the most part so reliable and smooth running, that – as with the processes of vision – they are really quite hard to think about. It is perhaps on account of this that one theme running throughout the chapter involves the importance of neuropsychological observations and studies in understanding the cognitive psychology of memory. Of course, it is the case that any memory impairment will itself be open to a variety of interpretations. Despite this, however, you will see in Chapter 8 that neuropsychological data have played an important part in the development of theories about the nature of memory.

In Chapter 9, 'Working memory', the focus of interest narrows further to take in just the memory stores and/or processes involved simply in maintaining whatever information an individual has in mind, or in executing whatever tasks they are engaged upon at a particular moment. Although there are necessarily considerable areas of overlap between the two chapters, the altered focus of interest results in a definite difference in emphasis. Where Chapter 8 dealt with issues concerning how information comes to be stored and retrieved (or not), the emphasis in Chapter 9 is more on the way in which available information is made use of. You will soon find, however, that while the focus of interest in Chapter 9 may seem rather narrower, working memory turns out to be an extensive topic in its own right.

As the chapter explains, the notion of working memory elaborates and extends upon the older and simpler idea of a short-term memory store. Working memory is conceived as a workspace with a limited capacity. But just as in Chapter 8 it proved necessary to postulate a variety of different kinds of memory, so it turns out that working memory itself fractionates into a number of component parts. Evidence for these separate components comes from studies employing various techniques for selectively interfering with cognitive performance. Once again neuropsychological data bear strongly upon the issues, and evidence is also adduced from studies employing neuroimaging techniques.

The history of the idea of working memory provides a good illustration of a point discussed in Chapter 1. You will see how the range of application of the theory of working memory has been extended as the theory has developed, and how with this extension researchers have become more confident of their theory. Chapter 9 also provides a discussion of the importance of computer modelling in the development and testing of cognitive theories, and introduces some illustrative examples. This chapter, therefore, previews a topic that is further expanded upon in Chapters 19 and 20.

One final theme to be found in Chapter 9 is that of individual differences. As described in Chapter 1, cognitive psychology as a whole tends to play down individual differences in favour of an emphasis on what it is that people have cognitively in common. This is similar to the way in which anatomists emphasize the considerable similarities in people's bodies ahead of their individual variations. But psychology, to an even greater degree than anatomy, cannot afford to overlook individuality for long. In Chapter 9 you will see how cognitive psychologists can make use of individual differences to test their theories, and also utilize their theories to explain individual differences in cognition.

This page intentionally left blank

CHAPTER 8

LONG-TERM MEMORY

Encoding to retrieval

Andrew Rutherford, Gerasimos Markopoulos, Davide Bruno, and Mirjam Brady-Van den Bos

1 INTRODUCTION

Everyone appreciates how useful it is to have a good memory. However, fewer people appreciate that having a good memory is not just useful – it is vital to the way we live our lives. Quite literally, our memory contains everything we know. Yet, despite the vast amount of information stored, memory almost always provides accurate and rapid access to the pertinent information we require. It is memory that tells us who we are and what we have done, it is memory that provides us with the words and grammar required to construct comprehensible sentences, and it is memory that holds the information that lets us recognize different types of cars, dogs, or sporting events, or make a cup of tea or coffee. Given the essential role of memory in our lives, it is not surprising that memory has been an active area of research in psychology since its first scientific investigation by the German philosopher Hermann Ebbinghaus in the 1880s.

This chapter focuses on **long-term memory**, particularly **episodic memory**, although there is also some mention of **semantic** and **implicit memory**. Episodic memory is a record of the episodes that constitute our lives. Episodic memory provides a description of what you have experienced (and thought) throughout your life. How these memories contribute to your understanding of yourself is considered in Chapter 17. This chapter presents accounts of how episodic memory operates and some pertinent experimental evidence. Researchers interested in normal memory usually examine people with normal memories, but they also examine people with abnormal memory resulting from physical damage to the brain. This may seem a peculiar way of finding out about normal memory, but an accurate account of normal memory operation should also explain why and how its manner of operation changes when damage is sustained. More formally, it can be said that data from neuropsychological studies provide useful constraints on psychological accounts of normal memory. Of course, such studies also provide beneficial insight into the memory problems experienced by brain-damaged people.

Logically, memory involves three stages, **encoding**, **storage**, and **retrieval** (getting information in, holding it, and then getting it back out). Typically, psychologists examine memory by presenting material and then observing what can be remembered. Different manipulations can be applied at the encoding, storage, and retrieval stages, depending on the purpose of the study. However, irrespective of whether encoding, storage, or retrieval is of interest, all stages will have been involved when information is remembered.

There are two major theoretical approaches apparent in memory research and in contemporary cognitive psychology generally: the processing approach and the systems approach (e.g. Foster and Jelic, 1999; Surprenant and Neath, 2009). As the names suggest, the processing approach emphasizes psychological processing, while the systems approach emphasizes structures or systems, usually with distinct brain locations. It will be apparent from the accounts presented in the following sections that while most researchers adopt a processing approach, they also are influenced by the systems approach, and the two approaches certainly are not incompatible. Nevertheless, in some sections, especially Section 3, where memory stores and systems are considered, both approaches compete to provide the superior theoretical account.

SUMMARY OF SECTION 1

- Our long-term memory contains all that we know and all that makes us who we are.
- Our memory operation is usually very efficient.
- Episodic memory is the record of our life experiences.
- Neuropsychological findings can constrain psychological accounts of normal memory.
- Memory involves three logical stages: encoding, storage, and retrieval.
- The processing and systems approaches are the major theoretical perspectives in memory research.

2 ENCODING

Encoding describes the way in which objects and events in the world come to be represented in memory. Our normal perception of objects and events requires considerable encoding, but further encoding processes can produce memory representations of objects and events that differ from those arising from perceptual processes alone.

2.1 Levels of processing

Craik and Lockhart's (1972) levels of processing (LOP) article exerted a large influence on memory research. Previously, accounts of memory performance were based on a fairly rigid memory systems viewpoint encapsulated by Atkinson and Shiffrin's (1968) '**multi-store**' memory model (see Chapter 9). This proposed separate sensory registers for each sense modality, along with a **short-term memory** store and a long-term memory store. Craik and Lockhart considered the multi-store model, but contrary to its proposals, they argued that the major determinant of an item's memorability was the LOP it received at encoding, and not the store in which it was held or how long it was held in this store (e.g. Baddeley *et al.*, 2009). Craik and Lockhart presumed that processing followed a fixed sequence of levels, from early perceptual processes through pattern recognition to the extraction of meaning. According to Craik and Lockhart, memories are not items specially constructed for storage, but are simply the traces of the processing operations left in the cognitive system. The greater the **depth of processing** applied to an item, the more elaborate, longer lasting, and stronger are the memory traces produced, and so the more likely it is that the item will be remembered (see Box 8.1).

Of course, processing need not continue through all levels. It may stop at any point when attention is diverted elsewhere, or, at any given level, the processing already engaged may simply repeat rather than proceed through further levels. An example of repetitive processing is verbally rehearsing a telephone number to keep it 'in mind' before calling the number. Craik and Lockhart labelled this **Type I processing**, while **Type II processing** was the label applied to processing that proceeded through further levels. (NB This is wholly different to the use of the terms Type 1 and Type 2 processing in relation to reasoning – see Chapter 12.)

BOX 8.1 RESEARCH STUDY Levels of processing (LOP) and memory performance

Craik and Tulving (1975, Experiment 1) presented subjects with a question followed by a word. Subjects had to answer 'yes' or 'no' to the question, and then later their recognition memory for the words was tested (see Table 8.1).

As the words were presented visually, visual processes were always engaged. Graphemic perceptual processing alone was required for question 1, but phonetic perceptual processing must also be engaged to answer question 2. Questions 3 and 4 both required semantically oriented processing, but more elaborative semantic processing was required to answer question 4 than question 3. As predicted, subjects' recognition memory performance increased with deeper LOP.

TABLE 8.1

Question	Yes	No
1 Is the word in capital letters?	TABLE	table
2 Does the word rhyme with 'weight'?	crate	market
3 Is the word a type of fish?	shark	heaven
4 Does the word fit in this sentence? 'The man peeled the ______.'	orange	roof

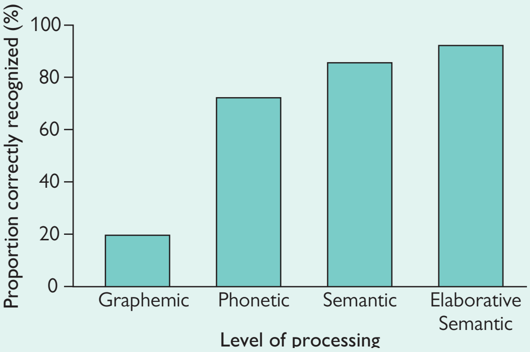


FIGURE 8.1 Recognition as a function of LOP.

Although the LOP framework provided a different perspective on the nature of memory and had substantial empirical support, there are problems with the account (e.g. Baddeley, 1978). Obtaining a valid measure of processing depth that is independent of the amount remembered is a particular problem. Without this, processing level can be defined only in a circular fashion – deeper levels of processing improve memory performance, so any improved memory performance can be taken to indicate a deeper level of processing and any lack of memory performance improvement can be interpreted as a failure to deepen the LOP at encoding. The consequence of such a circular definition is that LOP predictions cannot be refuted and so cannot be tested properly. Another problem is the assumption of a fixed sequence of processing levels from perception to semantic elaboration. Although this view is in keeping with the serial processing perspective of the 1970s, there is greater appreciation now of how much parallel processing underlies cognition.

There are also memory phenomena for which the LOP framework offers no explanation. For example, Glenberg *et al.* (1977) and Rundus (1977) presented numbers for subjects to remember, but to stop them rehearsing the numbers they had to rehearse words for various durations. In these circumstances, it was expected that the words would be maintenance rehearsed. However, subjects were asked to **free recall** the words at test, rather than the numbers. As predicted by Craik and Lockhart's LOP account,

BOX 8.2 RESEARCH STUDY Distinctive processing and level of processing

Subjects were presented with nouns they had to process in a semantically distinct (S–D) fashion by providing a descriptor (e.g. an adjective) that would be used infrequently to modify the noun (e.g. thin house). Semantically non-distinct processing (S–ND) was fostered by having subjects provide a frequently used descriptor to modify the noun (e.g. large house). Phonetically distinct processing (P–D) was achieved by presenting subjects with nouns pronounced differently to the way their spelling suggests, but subjects had to pronounce the words in line with their spelling (e.g. the usually silent 'b' in comb had to be pronounced at the end of the word). Phonetically non-distinct processing (P–ND) of nouns was obtained by having subjects maintain the conventional pronunciation.

The results showed that there was very little difference in recognition performance after semantic distinctive, semantic non-distinctive, and phonetic distinctive processing, but there was a significant drop after phonetic non-distinctive processing. Therefore, semantic processing enhances memory performance, but distinctive processing, even with phonetic processing, can lift memory performance to the level observed with semantic processing.

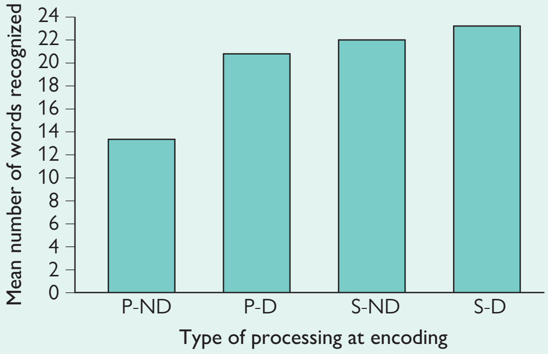


FIGURE 8.2 Correct recognition as a function of experimental conditions.

additional time spent maintenance processing the words did not benefit memory, as measured by free recall (Rundus, 1977). However, while Glenberg *et al.* (1977) observed the same with free recall, they also found that additional maintenance rehearsal improved **recognition memory**. LOP provides no account of the benefit recognition memory obtains from maintenance rehearsal, not least because the LOP framework focuses on encoding operations and pays no heed to retrieval operations. Later, we shall see how memory models have developed to account for such findings.

It has also been known for some time that distinctive items are well remembered (e.g. Koffka, 1935). According to LOP, **distinctiveness** results from greater depth of processing and semantic elaboration (e.g. Lockhart *et al.*, 1976). However, Eysenck and Eysenck (1980) manipulated distinctive processing independently of LOP (distinctive processing emphasizes differences between stimulus items) and observed that it exerted a beneficial effect on memory performance in addition to the effect of LOP (see Box 8.2). The issue of distinctive encoding is addressed in the following sections and again in Section 4.2.

2.2 Item specific and relational processing

Mandler (1979) provides a useful illustration of item-specific and relational processing – what he termed integrative and elaborative organization (see Figure 8.3). Item-specific processing binds the cognitive processes carrying the components underlying the mental representation, so enhancing their joint operation and coherence. One example is practising saying a word. This kind of item-specific processing benefits pronunciation fluency, and by analogy, item-specific processing is often described as enhancing processing fluency. Another consequence of item-specific processing is that item representations are likely to become increasingly differentiated – the nature of the fluency of processing of item A will become increasingly different from the nature of the fluency of processing of item B, and so item-specific processing underlies distinctiveness (Hunt and McDaniel, 1993). In contrast, relational processing establishes connections between different entity representations. For example, seeing a cat and imagining it being chased by a dog would establish a relation (i.e. chasing) between the two entity representations (cat and dog). According to Mandler (1979), maintenance rehearsal results in integration (i.e. employs item-specific processing), while semantic processing results in elaboration (i.e. employs relational processing).

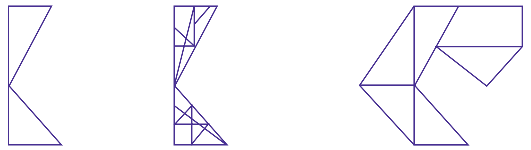

(a) Item (b) Item after integration (c) Item after elaboration

FIGURE 8.3 A graphic analogy of integration (due to item-specific processing) and elaboration (due to relational processing) adapted from Mandler (1979). Item-specific processing enhances the coherence of the cognitive processes carrying the mental representation components (depicted by the links between the components of the representation). Relational processing establishes connections between the mental representations of the target item and other items.

Hunt and Einstein (1981) examined the relationship between item-specific and relational processing, and free recall and recognition (see Box 8.3). They found that free recall and recognition were greater after relational processing followed by item-specific processing than after relational processing followed by more relational processing. They also found that free recall was greater after item-specific processing followed by relational processing than after item-specific processing followed by more item-specific processing, indicating that free recall benefits from both relational and item-specific processing.

However, similar high levels of recognition performance were observed when item-specific processing was followed by relational processing or more item-specific processing. Although ceiling effects masking recognition performance differences is a mild concern, these findings suggest recognition obtains greater benefit from further item-specific processing than free recall.

BOX 8.3 RESEARCH STUDY Item-specific and relational processing, free recall and recognition

Hunt and Einstein (1981, Experiment 1) presented subjects with either a categorized list of 36 words (six words from each of six categories) or 36 unrelated words. Subjects were assumed to process the categorized words spontaneously in a relational fashion, while subjects receiving the unrelated words were assumed to process them spontaneously in an item-specific fashion. For both categorized and unrelated lists, free recall and recognition were tested. However, prior to these tests, subjects were required either to sort the words into specified categories (a relational processing task) or to rate the pleasantness of the words (an item-specific processing task). Subjects read a short story for one minute before trying to free recall the 36 words. Recognition was tested after free recall. Table 8.2 below presents the average free recall and recognition scores.

TABLE 8.2

Presentation	**Categorized list**		**Unrelated list**	
	Relational processing (RP)		Item-specific processing (ISP)	
Task	Categorization (RP)	Pleasantness (ISP)	Categorization (RP)	Pleasantness (ISP)
Free recall[1]	.42	.48	.47	.33
Recognition[2]	.73	.93	.89	.91

[1] *Correct free recall as a proportion of total number of items presented (i.e. 36).*

[2] *AG scores – a non-parametric measure of recognition sensitivity.*

2.3 Other forms of encoding processing

A great variety of tasks have been employed to manipulate the encoding processing of stimulus materials. Craik and Lockhart's (1972) theoretical perspective emphasized processing depth (perceptual to elaborative semantic processing), while the distinction between item-specific and relational processing emphasizes the processing focus (single to multiple items). In principle, it is possible to combine these different forms of processing, but relational processing is almost always semantic. For example, relational semantic processing can be obtained by having subjects construct a story from the stimulus words presented (e.g. Rutherford, 2000), whereas the common request for subjects to rate the pleasantness of stimulus items (e.g. Hunt and Einstein, 1981) or state whether each word refers to a natural or manmade object (e.g. Markopoulos *et al.*, 2010) promotes a low level of item-specific semantic processing – perhaps better described as non-relational semantic processing. Asking subjects to keep a tally of the number of stimulus words containing three or more vowels can promote lower level (graphemic or phonetic) item-specific processing (e.g. Rutherford, 2000).

Typically, people engage semantic processing when asked to memorize material (e.g. Bower, 1972; Craik, 1977). However, Matzen and Benjamin (2009) argue that study circumstances affect subjects' encoding aims and so the way material is processed at encoding. For example, subjects presented with *words* to study for a later memory test engaged more perceptually-oriented processing, as at recognition they were likely to be misled by the perceptual similarities of stimulus items (e.g. recognizing *blackbird* rather than *blackberry*). In contrast, subjects presented with *sentences* to memorize engaged more semantic processing and so were much more likely to make semantic errors (e.g. recognizing *sleep* instead of *slumber*).

Another very effective technique for encoding information into long-term memory, which affects

BOX 8.4 RESEARCH STUDY Self memory bias created through ownership

Van den Bos *et al.* (2010) created a self memory bias via ownership. Subjects had to imagine they and a fictive other student, 'John', had won a basket of supermarket shopping items. Two baskets were shown on a monitor and subjects were told which was theirs. Shopping items were then shown on the monitor and subjects had to put these into their own or John's basket on the basis of the colour border around the item. This basic sorting task did not require subjects to engage in elaboration or self evaluation.

Performance on a remember (R)-know (K) recognition task (see Section 4.4) was better for self owned than for other owned items but only with R responses. This suggests elaboration of item representations occurs spontaneously when perceiving a self cue – that is, when the self is involved in encoding, material is encoded such that it is easily retrieved in great detail, making self referential encoding a very effective encoding technique.

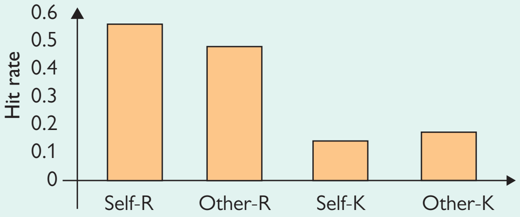


FIGURE 8.4 Hit rate as a function of ownership and response type.

recall and recognition, is linking information to the self. (See Box 8.4 and Symons and Johnson, 1997, for a review of the self reference effect.)

SUMMARY OF SECTION 2

- The levels of processing (LOP) framework was a counter to the multi-store memory model.
- The levels of processing framework asserted that memorability was due to the level of processing engaged at encoding – not the store in which an item was held.
- Distinctive (item-specific) processing can benefit memory independently of the level of processing.
- Relational and item-specific are two important types of processing.
- Item-specific processing enhances the operation and coherence of the cognitive processes carrying mental representations, and benefits processing fluency.
- Relational processing establishes connections between item representations.
- The self reference effect (SRE) refers to superior recall or recognition following self referencing than following other types of encoding.
- The SRE is apparent in remember (R) but not in know (K) responding.

3 MEMORY STORES AND SYSTEMS

A memory store is where non-active memory representations are held. For example, imagine your favourite item of clothing. When not in use, the memory representation upon which this image depends will be held in a memory store. Memory systems include memory stores, but memory systems also include all the processes that operate when memory representations are active, such as the processes that generate the image of your favourite item of clothing. The memory systems perspective is that memory stores and memory processing are localized in the same part of the brain. This view receives some support from research on connectionist systems, where representation and processing are intimately related.

Neuropsychological and neuroimaging data have exerted greater influence on theoretical accounts of memory systems than on theoretical accounts of memory encoding or memory retrieval. However, there also has been considerable theoretical and empirical dispute about memory systems accounts.

3.1 Declarative and non-declarative memory systems

An influential memory systems perspective was presented by Squire (e.g. Cohen and Squire, 1980; Squire, 2007). Squire's main focus was the amnesic syndrome (see Box 8.5), but his memory systems perspective also addressed and was soon applied to normal memory. The amnesic syndrome is a gross memory impairment due to an organic cause.

However, amnesics also exhibit normal or close to normal learning on tasks such as priming and psychomotor skills (for a brief review see Eysenck and Keane,

BOX 8.5 CASE STUDY HM and the amnesic syndrome

Henry Gustav Molaison died at the age of 82 in December 2008. To protect his privacy prior to his death he had been known as HM, and he was probably the most studied individual in the history of psychology.

When he was 27, HM underwent brain surgery in an attempt to treat his intractable epilepsy and those parts of his brain considered to be the origin of his epileptic seizures were removed (see Figure 8.5). The operation was successful to the extent that his epilepsy could now be controlled by drugs, but a tragic and unforeseen outcome was that HM became profoundly amnesic (Scoville and Milner, 1957; Milner, 1966). Although HM was unable to remember most events that occurred in the couple of years prior to his operation (retrograde amnesia, see Figure 8.6), he seemed to have a pretty normal memory for events that occurred prior to this time. However, most strikingly, HM seemed to have lost most of his ability to form any new memories (anterograde amnesia, see Figure 8.6). The removal of the anterior two-thirds of the hippocampus bilaterally (i.e. from both sides of his brain) is thought to have been responsible for his amnesia (e.g. Squire, 1987).

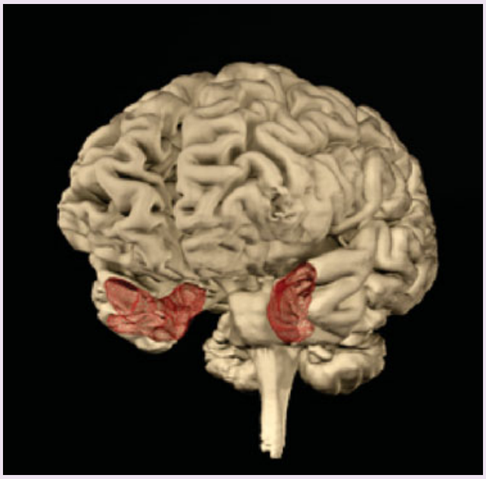


FIGURE 8.5 A three-dimensional reconstruction of HM's brain (http://thebrain observatory.ucsd.edu/content/3d-reconstruction-brain-mri-data). HM's brain was lesioned bilaterally (on both sides), with each lesion extending 5 cm from front to back. HM lost about two-thirds of his hippocampus, parahippocampal gyrus, and amygdala, as well as some of his anterolateral temporal cortex. However, as all of HM's entorhinal cortex (which forms the major sensory input to the hippocampus) was removed and the remaining 2 cm of hippocampal tissue appeared atrophic, it is likely that his hippocampus may have been completely non-functional.

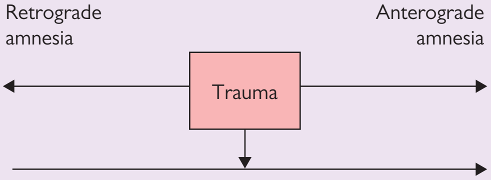


FIGURE 8.6 An illustration of the retrograde and anterograde amnesia.

HM stayed with his parents for some time after his operation, but as his memory problems made it impossible for him to live without supervision, he moved to a nursing home in 1980. HM's father died in 1967 and his mother in 1977, but in 1986, HM thought he still lived with his mother and was unsure if his father was alive (Parkin, 1993). HM could read the same book or magazine repeatedly without any recollection of having read it before and, typically, after spending all morning with psychologists doing various tests, he could not remember the testing session or recognize the psychologists when they returned in the afternoon. Like others suffering from the amnesic syndrome, HM was able to interact and converse quite normally despite his gross memory impairment. He retained a normal immediate memory span and demonstrated memory for a variety of perceptual and motor tasks, although he had no memory of the learning episodes underlying his perceptual and motor memory performance.

The amnesic syndrome seems to manifest whenever there is bilateral hippocampal damage. There may be a variety of different reasons for such damage, but most amnesics suffer from Korsakoff's syndrome due to chronic alcoholism. A thiamine deficiency is one consequence of chronic alcoholism and results in damage to parts of the brain, especially within the diencephalon (e.g. hippocampus and amygdala) and frequently the frontal lobes.

2010; for a more detailed account see Spiers *et al.*, 2001). Squire argued that the tasks on which amnesics exhibited their gross impairment depended upon one particular neurologically-based memory system, whereas those tasks amnesics still could perform relied upon a different neurologically-based memory system. Initially, Squire applied a distinction between declarative and procedural knowledge drawn by the philosopher Ryle (1949). Squire later modified his classificatory scheme (e.g. Squire, 1987; Squire, 1992) to distinguish between a single declarative memory system and a heterogeneous group of non-declarative memory systems (see Figure 8.7).

Declarative memory refers to memory for facts. Initially, Squire described declarative memory as being memory for certain process outputs, but later the key feature of the declarative memory system was described as being conscious awareness of the information retrieved (cf. explicit memory). A declarative format provides an explicit representation of information, which presumably benefits conscious access. However, so little is known about consciousness and its relationship with representational format that equating declarative memory with conscious access should be appreciated as a change in the nature of the classification rather than an expression of any understanding. Declarative memory would operate if you were asked 'What is the capital of Scotland?' The name of the city, 'Edinburgh', should enter consciousness and you would be able to state 'The capital of Scotland is Edinburgh'. Similarly, if asked 'Was this picture presented before?', a memory of the previous presentation of the particular picture should enter consciousness and, again, it would be possible to state 'This picture was presented before'. In contrast, non-declarative memory refers to a group of behaviours that are affected by stored information (i.e. memory), but there is no conscious awareness of the information retrieved (cf. implicit memory). Unfortunately, while these descriptions may be consistent with experience, they provide no explanation of how and why these conscious and unconscious experiences arise.

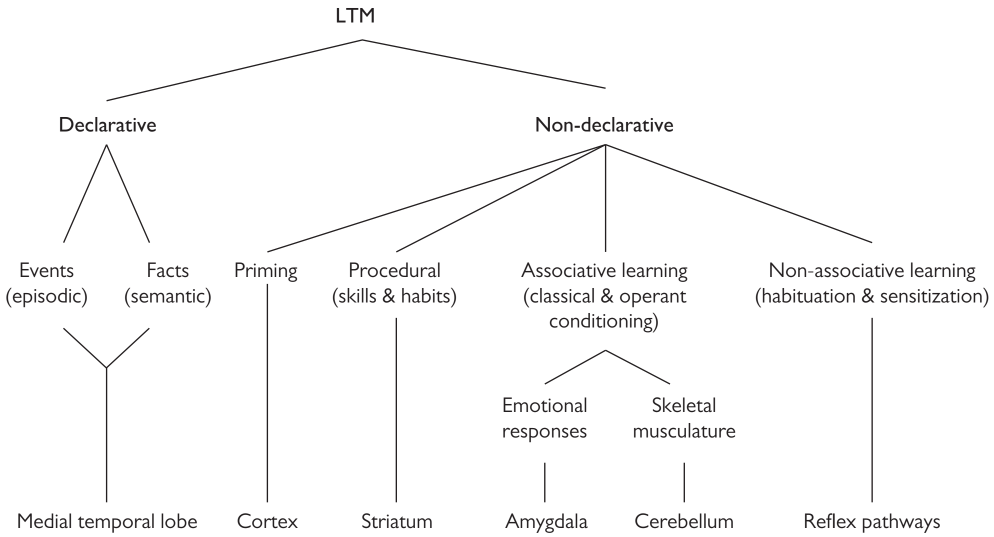

LTM
Declarative
Non-declarative
Events (episodic)
Facts (semantic)
Priming
Procedural (skills & habits)
Associative learning (classical & operant conditioning)
Non-associative learning (habituation & sensitization)
Emotional responses
Skeletal musculature
Medial temporal lobe
Cortex
Striatum
Amygdala
Cerebellum
Reflex pathways

FIGURE 8.7 Declarative and non-declarative memory systems and the brain regions with which they are associated. Adapted from Squire (2007). NB Medial temporal lobe structures emerge as elaborations of the medial edge of the temporal cortex.

3.2 Multiple memory systems

Tulving and associates are strong advocates of a multiple memory systems perspective. (Table 8.3 presents the various systems and sub-systems of human learning and memory proposed by Schacter and Tulving, 1994, and Schacter *et al.*, 2000.) Note that Schacter and Tulving present four long-term memory (LTM) systems with nine sub-systems, but the primary or working memory system is a temporary memory system and not a LTM system. Discussion here concentrates on the LTM systems, particularly the distinction between episodic and semantic memory.

Semantic memory is our general knowledge store and contains all the information underlying our understanding of the world. For example, it provides the information we use to recognize or describe different types of animals, objects, etc., it provides the information for using and understanding language, and it stores the sort of information we would employ to choose our ideal summer holiday destination. The question 'What is the capital of Scotland?' would tax semantic memory. However, no personal experience accompanies information retrieved from semantic memory.

Episodic memory provides a record of a person's experiences. It stores information about the occurrences that make up a person's life along with contextual information about these events, such as information about where they occurred and the person's thoughts and feelings about the event. In fact, the episodic memory system can output such a range of different kinds of information about the event that episodic remembering has been described as re-experiencing the past and mental time travel (e.g. Baddeley *et al.*, 2009). The question 'Have you seen this picture before?' would tax episodic memory.

3.3 Comparing declarative and non-declarative, and multiple memory systems

An initial glance at the two systems presented in Figure 8.7 and Table 8.3 suggests considerable overlap, but this impression is misleading. The only true match is between Squire's non-declarative priming system and Schacter and Tulving's perceptual representation

TABLE 8.3 Schacter and Tulving's systems and subsystems of human learning and memory (Schacter and Tulving, 1994; Schacter *et al.*, 2000)

System	Other labels	Sub-systems	Retrieval type
Primary memory	Working memory	Visual Auditory	Explicit
Semantic	General Factual Knowledge	Spatial Relational	Implicit
Episodic	Personal Autobiographical Events		Explicit
Perceptual representation	Non-declarative	Visual word form Auditory word form Structural description	Implicit
Procedural	Non-declarative	Motor skills Cognitive skills Simple conditioning Simple associative learning	Implicit

system. Both Squire and Schacter and Tulving present procedural memory systems, but while Schacter and Tulving accommodate simple conditioning (i.e. classical but not operant conditioning) and simple associative learning within the procedural memory system, Squire attributes associative learning (i.e. classical and operant conditioning) to a separate non-declarative memory system. Comparing non-declarative memory systems also reveals that Schacter and Tulving omit any mention of operant conditioning and non-associative learning.

The same quick glance also suggests agreement between Squire's declarative (facts and events) memory system and Schacter and Tulving's semantic (facts) and episodic (events) memory systems. However, Squire presents a single declarative memory system responsible for the retrieval of both facts and events, whereas the first multiple memory systems distinction was drawn between semantic and episodic memory (Tulving, 1972). Squire's main claim is simply that declarative memory output enters consciousness. Schacter and Tulving agree, but they also attribute a separate and unique character to the type of consciousness supported by episodic memory.

The multiple memory systems perspective, particularly the semantic and episodic distinction, has received far greater criticism than Squire's account of declarative and non-declarative memory systems (e.g. McKoon *et al.*, 1986; Neely, 1989; Roediger *et al.*, 1999; Toth and Hunt, 1999; Surprenant and Neath, 2009; Shanks, 1997). However, a benefit of this has been greater theoretical development of the multiple memory systems account. For example, Roediger (e.g. Roediger *et al.*, 1990; Roediger *et al.*, 1999) criticized the multiple memory systems perspective for failing to define the criteria by which systems should be classified and distinguished. Clearly, it is essential to know exactly what a memory system is (and isn't) when it is presented as an explanation of memory phenomena. A theoretical account of what constitutes a memory system is vital because it defines (i.e. constrains) what is and what isn't a memory system and so distinguishes its predictions from other theoretical accounts. Roediger *et al.* (1999) attribute the proliferation of inconsistent memory system models to this failure. Crowder (1993) states that a memory system must involve more than just the identification of the brain structure(s) carrying out the cognitive operations for a specific task. As all cognitive processes have a neural basis, simply defining a memory system as the brain structure(s) that carry out the cognitive operations for a specific task does no more than state where in the brain these processes run and is as compatible with the processing perspective as it is with the multiple memory systems perspective. However, if more and more memory systems are postulated, then more and more brain locations will be running different cognitive processes. The processing perspective tends not to identify processes with brain locations, but many processes running in different parts of the brain is the default processing perspective. Therefore, as more and more memory systems are postulated, so the difference between multiple memory systems and processing perspectives diminishes.

In response to these and related criticisms, Schacter and Tulving (e.g. Schacter and Tulving, 1994; Schacter *et al.*, 2000) presented the following three criteria to define a memory system:

- *Class inclusion operations* – different memory systems operate on different kinds of information from different classes or domains.
- *Properties and relations* – separate memory systems are defined differently with respect to a set of properties and relations, such as their information class or domain (see above), system function and rules of system operation, and neural substrates.
- *Convergent dissociations* – memory systems should exhibit multiple and consistent dissociations, e.g. between two specific tasks, between sets of similar tasks, between patients with and without specific brain damage, or when the different memory systems have different physiological characteristics.

Nevertheless, Roediger *et al.* (1999) considered these definitions to be inadequate, and by way of illustration they made the point that these criteria would draw an untenable distinction between a recall memory system and a recognition memory system. A theoretically complete specification of what constitutes a memory system remains elusive.

3.3.1 The semantic–episodic distinction: cognitive/neuropsychological evidence

The semantic and episodic distinction is a useful descriptive heuristic, but many psychologists remain sceptical about the multiple memory systems perspective, and many doubt the specific distinction between semantic and episodic memory systems (e.g. Baddeley, 1984). As all memories are encountered first through experience, it seems the first destination of all encoded information must be episodic memory. Therefore, for semantic memory to develop there must be a well-worn route from episodic to semantic memory (see Box 8.6). However, most, if not all, of our episodic memories are understood in semantic terms. Anderson and Ross (1980) demonstrated that episodic memory information affects the retrieval of semantic information. Given the continual sharing and blending of semantic and episodic information required, a single system rather than two separate systems seems a more attractive and likely design. To address some of these criticisms, Tulving (1984) proposed that the easy influence of semantic information on episodic memory was due to episodic memory being a recent evolutionary development arising out of, but embedded within, semantic memory. Unfortunately, however, exactly how this enabled access and influence was not explained.

The memory systems perspective and the distinction between semantic and episodic memory was adopted readily by neuropsychologists. Neuropsychological data obtained from amnesic patients already had influenced views about short-term and long-term memory (e.g. Baddeley and Warrington, 1970) and it also exerted a strong influence on the semantic-episodic memory distinction.

The amnesic syndrome includes a retrograde component extending back months or even years from the amnesia-inducing trauma, but amnesics typically retain normal memory for events that occurred prior to this period (see Box 8.5). As the brain damage causing the amnesic trauma does not affect access to these long established memories, it suggests the damaged neural substrate is not the memory store. This, along with the apparent inability to form new memories, implies the amnesic syndrome is due to problems with one or more of the processes that encode or retrieve these new memories.

Tulving (1983) argued that the amnesic syndrome is due to a severe deficit in episodic memory combined with an intact semantic memory. The retention of amnesics' intellect and language skills demonstrates a substantial part of their semantic memory operates normally. Nevertheless, some studies indicated amnesics' semantic and episodic memory were equally impaired in acquiring new information, and the apparently normal operation of semantic memory arose from the use of semantic information acquired prior to the amnesic trauma (e.g. Cermak and O'Connor, 1983; Gabrieli *et al.*, 1988; Grossman, 1987). These findings appeared to contradict the view that the amnesic syndrome was due to a severe deficit

BOX 8.6 Does semantic memory arise from episodic memory?

Consistent with episodic memory being the first destination of all encoded information is a conception of semantic memory as an abstraction of episodic experience. Common aspects of episodes are, by definition, experienced repeatedly. In contrast, there is an inconsistent association between the common aspects of the episodes and the various contexts in which they occur. As a result, the common aspects of the episodes will be well learned and will be easily and speedily retrieved, while the associated contexts, without the benefit of such repetition, will become inaccessible or will fade from memory (e.g. Baddeley, 2002). A number of possible cognitive and biologically oriented systems have been described that, consistent with this perspective, utilize information retrieved (repeatedly) from episodic memory to construct semantic memory representations (e.g. Hintzman, 1986; McClelland *et al.*, 1995; Rolls and Kesner, 2006).

in episodic memory combined with an intact semantic memory.

However, Spiers *et al.* (2001) examined 147 cases of amnesia where hippocampus or fornix had been damaged. All cases exhibited episodic memory impairment, but most semantic memory problems were mild. Moreover, Vargha-Khadem *et al.* (1997) described the cases of three patients, two of whom had sustained bilateral hippocampal damage at a very young age – Beth at birth and Jon at 4 years old. Both exhibited extremely poor episodic memory, but they attended normal schools and their language development, literacy levels, and factual knowledge were within the normal range. (Beth and Jon's condition is known as **developmental amnesia** or **early onset amnesia**.) A few years later, Baddeley *et al.* (2001) reported the results of a number of assessments carried out with Jon between the ages of 19–22. Jon's level of intelligence was above average (WAIS-R full scale IQ = 114) and again, he exhibited normal performance in tests of language and general knowledge. However, as before, he showed impaired recall and appeared to lack the normal experience associated with episodic memory. Indeed, a subsequent study by Brandt *et al.* (2006) suggested that much of Jon's recall of episodic information was based on semantic memory, and so his episodic memory may have been even more impaired than originally thought. Verfaillie *et al.* (2000) also described a female amnesic, PS, who (somewhat uniquely) according to MRI scans had bilateral damage limited to the hippocampus. PS exhibited severely impaired episodic memory, but was able to perform normally on less demanding semantic memory tasks. Although PS did show performance impairment when the semantic task retrieval demands increased, these findings are generally consistent with Vargha-Khadem *et al.*'s (1997) suggestion that episodic memory relies upon the hippocampus, while semantic memory depends upon the entorhinal, perirhinal, and parahippocampal cortices, which are adjacent and input to (as well as receive output from) the hippocampus (also see Section 4.4 and Figure 8.14). As all of these structures are not only linked and adjacent, but occupy a relatively small area in the medial temporal lobe (MTL), it is likely that most amnesics (e.g. those studied by Cermak and O'Connor, Gabrieli, and Grossman) will have suffered damage to both hippocampus and adjacent areas, with the result that both semantic and episodic memory will be impaired. These findings are consistent with a distinction between semantic and episodic long-term memory systems, but the data from Beth, Jon, and PS also suggest that semantic memory may arise without episodic memory (cf. Tulving, 1984, and Box 8.6).

3.3.2 The semantic–episodic distinction: neuroimaging evidence

The multiple systems perspective and distinction between semantic and episodic memory also was adopted with some enthusiasm by researchers employing functional neuroimaging. One of the first things these new techniques revealed was that a large number of different regions of prefrontal cortex (PFC) were involved in semantic and episodic memory tasks. PFC involvement in memory has been attributed to executive control of memory processing (e.g. Cabeza and Nyberg, 2000), which is consistent with the problem solving activities (selecting, planning, organizing, and monitoring) underlying memory operation suggested by Norman and Bobrow's account of memory encoding and retrieval (Norman and Bobrow, 1979; Norman, 1982; see Section 4.2).

Functional neuroimaging research has observed activation in PFC and MTL as well as cerebellar regions during episodic memory encoding (e.g. Cabeza and Nyberg, 2000), while activation is observed in PFC and MTL as well as lateral parietal, anterior cingulate, occipital, and cerebellar regions during episodic memory retrieval (e.g. Cabeza and Nyberg, 2000). More recently, functional neuroimaging research has started to identify the types of processing implemented by specific areas of the hippocampal formation (e.g. Carr *et al.*, 2010).

Neuroimaging (and neuropsychological) research suggests that semantic memory is a collection of functionally and anatomically distinct systems. In these systems, information about each semantic attribute is usually linked to a sensorimotor modality (e.g. vision) and sometimes an even more specific property within that modality (e.g. colour; see Thompson-Schill *et al.*, 2006 and Patterson *et al.*, 2007 for neuroscience orientated reviews of semantic memory). Nevertheless, there are inconsistencies between the neuroimaging data, usually obtained from normal subjects, and the data obtained from the neuropsychological study of amnesic patients (e.g. Ryan *et al.*, 2008). For example, there is currently an absence of neuroimaging evidence supporting the special role of the entorhinal, perirhinal, and parahippocampal cortices in the development of semantic memory. Moreover, hippocampal activation has been associated with the retrieval of a number of different kinds of well-established semantic information, such as category exemplars (Ryan *et al.*, 2008), spatial relations between objects (Ryan *et al.*, 2010), famous faces (Bernard *et al.*, 2004; Leveroni *et al.*, 2000), historical events (Maguire and Mummery, 1999), and semantic associations (Henke *et al.*, 2003). Findings of this sort have maintained the debate about hippocampal involvement in semantic memory retrieval.

However, in a functional magnetic resonance imaging (fMRI) study, Hoscheidt *et al.* (2010) matched episodic and semantic tasks in terms of spatial content and observed hippocampal activation not only when episodic memories were retrieved, but also when semantic memories with spatial content were retrieved. They concluded that the hippocampus was involved in the retrieval of spatial information irrespective of its semantic or episodic designation, and so sharp dissociations between semantic and episodic memory systems must be overly simplistic. Prince *et al.* (2007) employed fMRI to investigate the relationship between episodic encoding and semantic retrieval. Using word stimuli, activation in left hippocampus (and thalamus) was associated only with episodic encoding, and activation in a posterior lateral region of the left temporal lobe was associated only with retrieval from semantic memory. In the left inferior PFC, a dorsal posterior region was active during retrieval from semantic memory, a mid-region was active during retrieval from semantic memory and episodic encoding, and a ventral anterior region was active during episodic encoding, but only when there was a high level of retrieval from semantic memory. Prince *et al.* suggest that these data indicate a unique role for the hippocampus in episodic encoding, and that the inferior PFC may have a complex role in the interaction between episodic encoding and semantic memory retrieval. Nevertheless, debate continues as to whether the amnesic syndrome is best described in terms of a single damaged declarative memory system, or a damaged episodic memory system and an intact semantic memory system (e.g. Bayley and Squire, 2005; Burianova and Grady, 2007; O'Kane *et al.*, 2004; Rajah and McIntosh, 2005; Squire *et al.*, 2004).

A constant theme of recent research on semantic and episodic memory is the issue of system interdependencies (even within semantic memory) with a strong emphasis on semantic and episodic memory systems operating in concert to support normal memory function (e.g. Baddeley et al., 2009; Eysenck and Keane, 2010). Greenberg and Verfaellie (2010) reviewed neuropsychological studies and concluded that semantic and episodic memory can influence each other at encoding and retrieval, and so theories should accommodate the interdependencies between the two types of memory. Indeed, the extent of these interdependencies has led researchers to suggest that a clear-cut distinction between semantic and episodic memory systems is too simplistic (e.g. Hoscheidt et al., 2010; Prince et al., 2007; Rajah and McIntosh, 2005). These issues also raise questions about exactly what constitutes a system, exactly what benefits accrue from the theoretically loaded notion of separate systems, and whether this is preferable to a process-orientated conception (e.g. Roediger et al., 1999). It seems that a theoretical interpretation of memory in terms of semantic and episodic systems still merits caution.

SUMMARY OF SECTION 3

- Non-active memory representations are held in a memory store.
- The memory systems perspective regards memory storage and processing as occurring within a system localized within the brain.
- Declarative and non-declarative memory systems and multiple memory systems have been proposed.
- The multiple memory systems perspective advocates a large number of memory systems and sub-systems, including episodic and semantic memory.
- A theoretically sufficient definition of what constitutes a system remains elusive.
- Debate continues as to which theoretical approach (processing or systems) or which memory systems perspective provides the best theoretical account of the data.
- Neuropsychological and neuroimaging evidence exert a strong influence on theory development in this area of research.
- If the multiple memory systems perspective is adopted, then substantial interaction between semantic and episodic memory systems must be accommodated.

4 RETRIEVAL

Retrieval is the label given to the way memory information is obtained for use. Retrieval involves finding, activating, and sometimes further processing pertinent memory representations.

4.1 Encoding specificity and transfer appropriate processing

Tulving and Osler (1968; see also Tulving, 1983) introduced the **encoding specificity** hypothesis to explain the role of cues in memory retrieval. They suggested that specific retrieval cues facilitate recall only if they and their relation to the target item are stored along with the target item. In their study, subjects were presented with target words written in capitals. Also presented with each target word were zero, one, or two weakly associated words written in lowercase (e.g. MUTTON, fat, leg; CITY, dirty, village). Subjects were told that the words in lowercase might help them remember the capitalized target words and to try and think about how the lowercase words were related to the target words. Tulving and Osler found a single weak associated aided recall of the target word, provided the weak associate had been presented at learning. Neither one nor two weak associates aided recall if they had not been presented at learning – recall was not assisted by the provision of these cues at test alone. Successful retrieval of the target item increased with the overlap between the information stored in memory and the information employed at retrieval (Tulving, 1979).

The **transfer appropriate processing (TAP)** hypothesis also emphasizes the overlap between encoding and retrieval, but Morris *et al.* (1977; see also Bransford *et al.*, 1979) presented TAP to assert the role of retrieval processing, which had been ignored in Craik and Lockhart's (1972) LOP account. TAP refers to the overlap between the processes engaged at encoding and retrieval and predicts best memory performance will be observed when the processes engaged at encoding transfer appropriately to retrieval (see Box 8.7). However, Craik (2002) noted that although Morris *et al.*'s (1977) data reveal the interaction between encoding and retrieval predicted by TAP, best memory performance was observed when semantic processing was employed at both encoding and retrieval. So, while the encoding–retrieval processing match is important, it may not be the main reason for the memory performance observed (see Section 4.2).

Encoding specificity deals with information and TAP deals with processing, but these distinctions may

BOX 8.7 RESEARCH STUDY Transfer appropriate processing

Morris *et al.* (1977, Experiment 2) presented subjects with a list of words (e.g. CAT, TABLE). All subjects received two types of orienting question: 'Does the word rhyme with hat? Does the word rhyme with label?', which require phonetic processing to answer, or 'Is it an animal? Do you sit at it?', which require semantic processing to answer. Subjects then split into immediate test and test after 24 hour groups. For each test group, half of the subjects were asked to recognize which of a set of words rhymed with any of the words presented originally (e.g. which of these words rhyme with those presented originally: FIRE, MAT, STAIR, CABLE, PAPER, etc.). The other subjects were given a standard, semantically-orientated recognition test (e.g. which of these words were presented previously? CAT, ROAD, POUND, TABLE, BALL, etc). Similarly, half of the semantic-orienting condition subjects received the rhyme test, while the others received a standard, semantically-oriented recognition test. Figure 8.8 presents the mean adjusted proportion of correctly recognized words as a function of orienting task and recognition task for the 24 hour delay test group, and Figure 8.9 presents a schematic of the TAP experimental design.

When rhyme (phonetic) processing was employed at encoding, memory performance was best when rhyme (phonetic) processing was employed at test. Similarly, when semantic processing was employed at encoding, memory performance was best when semantic processing was employed at test. So, as TAP predicts, memory performance is best when the encoding processes transfer appropriately to retrieval.

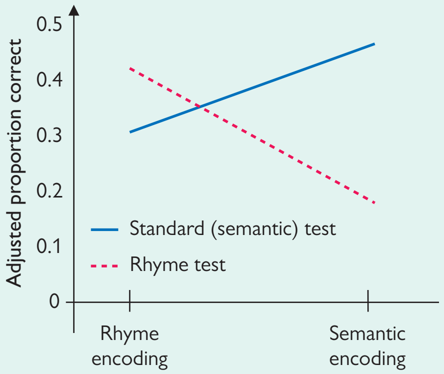


FIGURE 8.8 Adjusted proportion of items correctly recognized as a function of encoding task and test type after 24 hours.

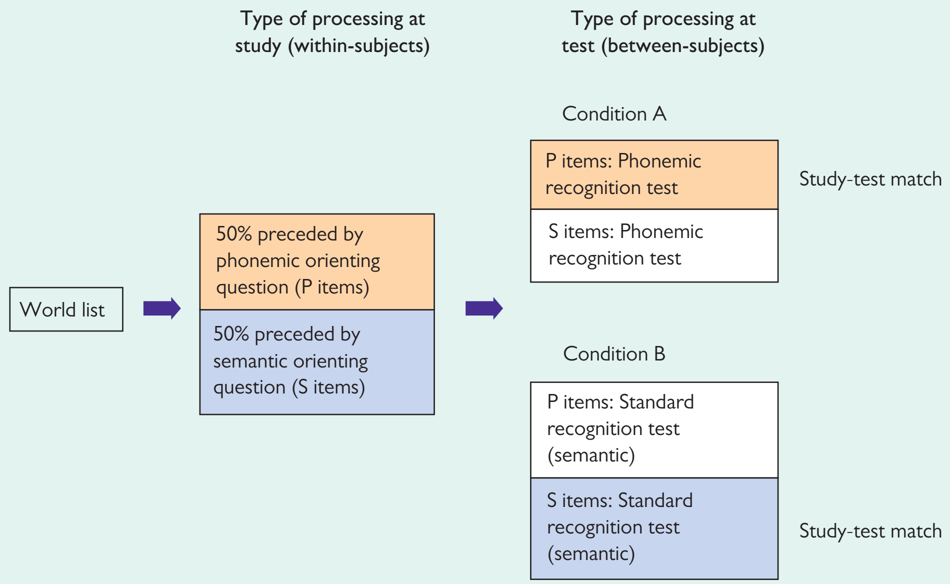

Type of processing at study (within-subjects)
Type of processing at test (between-subjects)
World list
50% preceded by phonemic orienting question (P items)
50% preceded by semantic orienting question (S items)
Condition A
P items: Phonemic recognition test
S items: Phonemic recognition test
Study-test match
Condition B
P items: Standard recognition test (semantic)
S items: Standard recognition test (semantic)
Study-test match

FIGURE 8.9 A schematic depiction of the TAP experimental design. *Source: based on Morris et al., 1977*

be different sides of the same coin. Both accounts emphasize the relationship between encoding and retrieval, and the benefit to memory performance when encoding conditions are repeated at retrieval. As information at encoding and at retrieval is manifest by cognitive processes, it may be more a matter of expression rather than psychological substance whether information or processes are repeated.

4.2 The diagnostic value of retrieval cues

Nairne (2002) argued that contrary to encoding specificity and TAP accounts, the match between the cues employed at retrieval and the information represented in memory may not be the key factor determining successful memory retrieval. Instead, Nairne argued that successful memory retrieval is determined by the diagnostic value of the retrieval cues employed – diagnostically valuable cues distinguish target memory records from all other records in memory. Indeed, Surprenant and Neath (2009) consider the 'relative distinctiveness' of a target item at retrieval to be a principle of memory (e.g. a single long word in a list of short words).

In a classic article, Norman and Bobrow (1979) suggested that entities may be encoded with varying degrees of specificity depending on a number of factors, such as task demands. Similarly, successful memory retrieval is determined by two factors: constructability – the extent to which pertinent descriptions of memory targets can be constructed – and discriminability – the extent to which target memory records can be discriminated from all other records in memory. Therefore, successful memory retrieval requires the construction of unique memory records that can be discriminated from all other records in memory. Unfortunately, however, even when a unique memory record is constructed, it may not remain unique after further items and events are encoded.

Table 8.4 presents the four retrieval aspects or subprocesses described by Norman and Bobrow (1979; Norman, 1982, added a fourth in an updated account). The particular circumstances and constraints under

TABLE 8.4 Four retrieval sub-processes with outlines of their inputs and outputs

Sub-process	Input to sub-process	Output from sub-process
Retrieval specification	(Partial) perceptual descriptions	Target description*
	(Partial) description(s) of purposes/needs	Verification criteria*
Matching	Target descriptions	Memory records
Evaluation	Memory records	Success or failure:*
	Verification criteria	if success terminate
		if failure diagnose
Failure diagnosis	Information from evaluation process	Revised retrieval specification

** – may require memory retrieval*

Source: Modified from Norman, 1982

which retrieval occurs are likely to influence the four sub-processes differently (e.g. time-limited recognition is likely to employ restricted retrieval specifications and evaluations, and abandon failure diagnosis). Norman and Bobrow's notions of target descriptions (i.e. memory probes), continuous cycles of retrieval specification, and matching, evaluation, and failure diagnosis (frequently based on partial information) running until retrieval succeeds or memory retrieval is terminated, were used and developed in many formal and informal memory models (e.g. Eich, 1985; Gillund and Shiffrin, 1984; Hintzman, 1988; Humphreys *et al.*, 1989; Morton *et al.*, 1985; Murdock, 1993) and are now part of the general vocabulary of memory operation.

Norman and Bobrow (1979) suggested that semantic encoding provides a very rich encoding 'vocabulary' in comparison with other coding schemes (e.g. there are a limited number of different phonemes, but the range of semantic elements and their combination is limited only by imagination). The greater distinctiveness (diagnostic value) available from such coding could explain why superior memory performance is generally observed when semantic encoding and retrieval are employed. It is worth noting that Norman and Bobrow's account reveals that a substantial part of what is classed as memory retrieval might be characterized better as problem solving (see Chapter 10). The retrieval specification, evaluation, and failure diagnosis sub-processes all involve some form of problem solving, assessment, or decision making with regard to the information to be input or output, and only the matching sub-process, which attempts to fit the target description to the memory records, can be regarded as a pure memory process.

4.3 Accounts of recognition memory

4.3.1 The single process signal detection model of recognition

Signal detection theory (e.g. Swets, 1959; Green and Swets, 1966; Macmillan and Creelman, 2005; Wickens, 2002) applied to memory provides the basis for the single process account of recognition memory performance (e.g. Ratcliffe *et al.*, 1992). Originally, signal detection theory (SDT) was a theory about how sensory and perceptual detection decisions were made – only later was it applied to recognition decisions (e.g. Banks, 1970; Lockhart and Murdock, 1970; Parks, 1966).

The single process signal detection (SPSD) model of recognition assumes that all of the events or items to be discriminated differ only in terms of the amplitude of a unidimensional quantitative measure. A great deal of recognition memory research labels this measure *familiarity*, but to avoid confusion here, *strength of evidence* will be used.

When applied to recognition memory, the SPSD model assumes that some retrieval process interrogates the memory representation and outputs a strength of evidence signal. Target events or items presented and encoded in an experiment will provide greater strength of evidence signals than other entities in memory that were not presented and encoded in the experiment. In principle, the process outputting the strength of evidence signal could assess and provide strength of evidence signals for all items in memory. However, the recognition test only requires discrimination between items previously presented (i.e. targets) and items not presented previously (i.e. distractors), so the strength of evidence signals for only these items need be considered. The SPSD model assumes that if strength of evidence signals for all targets and distractors were obtained and the frequency of these strength of evidence signals were plotted, then two distributions would be observed. Target items would be distributed higher (more to the right) on the strength of evidence axis than the distractor items that were not encoded previously in the experiment (see Figure 8.10).

When the x-axis represents 'strength of evidence', it is assumed that subjects are not aware of the strength of evidence signal distributions depicted in Figure 8.10. (Log likelihood recognition decision models do assume that subjects have some knowledge of these distributions, e.g. Shiffrin and Steyvers, 1997; McClelland and Chappel, 1998.) For each target or distractor item presented for recognition, the retrieval process outputs a strength of evidence signal and subjects have to make a recognition decision about each item based on each strength of evidence signal. Consequently, strength of evidence SPSD models assume that subjects make their recognition decisions depending on whether the item's strength of evidence signal is greater than a (response) criterion they set. All items returning (via the retrieval process) strength of evidence signals above the response criterion elicit 'recognized' responses and all items returning strength of evidence signals below the response criterion elicit 'not recognized' responses. Applying some mathematics relating the underlying target and distractor strength of evidence distributions, the proportion of correctly identified targets (i.e. hits) and the proportion of falsely recognized distractors (i.e. false alarms) provides subjects' responses in SDT terms. For example, the ability to discriminate between the distributions of targets and distractors – as assessed by the distance between two equivalent points on the target and distractor distributions (e.g. the apex of each distribution) – is termed d′ ('dee prime'), while the location of the response criterion (measured as a positive or negative distance from the point of intersection of the two distributions) is termed C ('see'). It is assumed that d′ is fixed but subjects can change C. (In Figure 8.10, the response criterion, C, is set at the point of intersection of the two distributions – the optimum point for maximizing HR and minimizing FAR.) A useful way of depicting both SDT expectations and subjects' actual responses is to plot HR against FAR for different response criteria. Usually, different response criteria are obtained by asking subjects to rate their confidence that each memory probe is a target. This form of plot is known as a receiver operating characteristic (ROC) curve. A ROC curve that is symmetrical along the minor diagonal (from top left corner to bottom right corner) is predicted by the classic single process SDT account (see Figure 8.11). The leftmost point on the ROC curve reflects the HR and FAR for the strictest criterion location (i.e. greatest confidence that the test item is a target). The sensitivity to targets and distractors is the same for each point on the ROC curve (an alternative name for a ROC curve is *isosensitivity* curve). Greater sensitivity is indicated by curves that fall towards the upper left region of the graph, where the hit rate is greater and the false alarm rate is lower. More liberal response criteria are indicated by points toward the right-hand ends of the ROCs, where both hit and false alarm rates are greater.

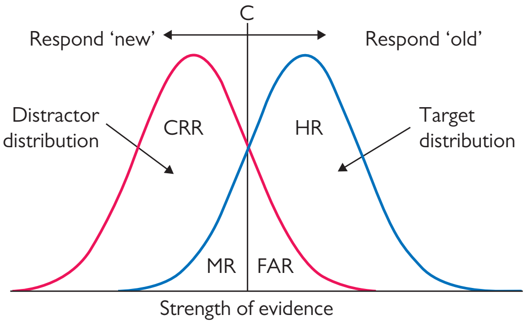


FIGURE 8.10 Signal detection model for recognition. HR = hit rate, FAR = false alarm rate, MR = miss rate, CRR = correct rejection rate, C = old/new criterion.
Source: Bruno et al., 2009

Unfortunately, however, the empirical data from recognition experiments most frequently provide asymmetrical ROCs. SPSD model accounts attribute

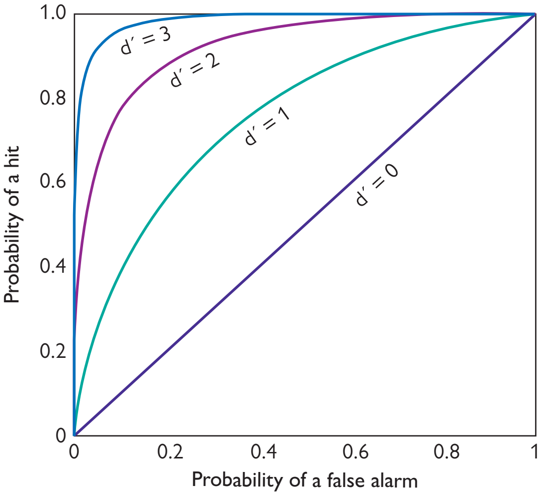


FIGURE 8.11 ROC curves. The curves falling towards the upper left region of the graph depict increasingly greater discrimination, as revealed by their greater d′ estimates. *Source: adapted from Macmillan and Creelman, 2005*

this discrepancy to the greater variation in target item strength of evidence distributions than distractor item strength of evidence distributions. Usually, the empirically determined ratio of target to distractor variances is 1:0.75 (e.g. Mickes *et al.*, 2007; Ratcliffe *et al.*, 1992). When the inequality of the target and distractor variances is accommodated, the single process model is termed the unequal variance model or the unequal variance SPSD. Although the unequal variance single process model maintains a single retrieval process, accommodating unequal target and distractor variances means the SPSD (like the DPSD – see next section) is a two factor model.

4.3.2 Yonelinas's dual process signal detection theory

Dual process accounts of recognition have a long history in psychology (e.g. Mandler, 2008; Yonelinas, 2002). All these accounts share the view that recognition can enlist two retrieval processes. A number of dual process accounts have been presented recently, but the most influential currently is the dual process signal detection theory (DPSD) presented by Yonelinas (e.g. Yonelinas 1994, 1997, 2002; Parks and Yonelinas, 2008).

In common with Atkinson and Juola (1973, 1974) and Mandler (1980), Yonelinas proposes that one retrieval process provides an assessment of the familiarity of items and events, while the second retrieval process is similar to that underlying free recall. Yonelinas suggests that these two independent retrieval processes are initiated simultaneously. A swift familiarity-based retrieval process assesses the fluency with which the items are processed: the more fluently an item is processed, the more familiar it feels (e.g. Jacoby and Dallas, 1981). As item-specific processing facilitates processing fluency, it should benefit familiarity-based retrieval too (Mandler, 1980). Yonelinas considers the familiarity-based retrieval process to be described well by the classic equal variance SDT. SDT assumes target and distractor distributions are continuous and their tails approach but never reach the zero point on the frequency axis. This means that there are always target and distractor items with memory strengths above zero, and so the familiarity-based retrieval process always outputs a continuous (i.e. not discrete) quantitative strength of evidence signal greater than zero. Therefore, familiarity-based retrieval output is predicted to be a classic SDT ROC curve, which is symmetrical along the minor diagonal.

Yonelinas labels the second retrieval process underlying recognition as recollection. In contrast to familiarity-based retrieval, recollection takes longer to complete and benefits from relational processing at encoding – the connections between the memory representations provide a variety of different routes (i.e. cues) to the target memory representation. Recollection is also conceived as a threshold process. When the retrieval cue information achieves the retrieval threshold, all of the qualitative information held by the particular memory record is released. However, if the threshold is not achieved, then zero qualitative information about an event is retrieved. Therefore, the recollection output is discrete – it results in remembering (various amounts) or it results in zero remembering. There is never a SDT situation where recollection retrieves some information, but the subject denies remembering. Consequently, recollection predicts a straight line ROC, which is asymmetric along the minor diagonal.

Yonelinas (1994) argues that asymmetric curvilinear recognition ROC curves are due to some items being recognized on the basis of recollective retrieval and some items being recognized on the basis of familiarity-based retrieval. In a number of studies, Yonelinas (1994, 1997, 2001) obtained data that fit his DPSD model predictions closely (see Box 8.8). However, discriminating between the different ROC curves predicted by SPSD and DPSD models can be a difficult task (e.g. Mickes *et al.*, 2010).

The DPSD recognition account and Hunt and Einstein's (1981) experimental results offer an explanation for the findings obtained by Glenberg *et al.* (1977) and Rundus (1977) reported in Section 2.1.

BOX 8.8 RESEARCH STUDY Empirical evidence of recollection and familiarity

Yonelinas (1997) proposed that recognition memory for single items (e.g. words) is based upon two independent processes: a familiarity-based retrieval process well described by an equal variance signal detection model, and a recollective retrieval process described by a threshold model. Signal detection models generate curvilinear ROCs, while threshold models generate linear ROCs.

Yonelinas provided further support for the DPSD account by comparing item recognition (e.g. single words and pictures) with pair recognition. All of the individual items presented for pair recognition had been presented before, but while some pairings were identical to what had been shown before, the other item pairings were new. Therefore, in contrast to item recognition, where brand new items were presented, the only way to recognize item pairs correctly was to access information in the memory about the association between the items presented at recognition, and only recollection provides this. If Yonelinas's claims are correct and recollection is a threshold process, then pair recognition should generate linear ROCs (asymmetric along the minor diagonal), while item recognition should generate curvilinear ROCs (symmetric along the minor diagonal).

Yonelinas (1997, Experiment 1) asked 20 subjects to engage in a mixed memory test where half of the test trials required single item recognition and half required pair recognition. When presented in the form of ROCs, as shown in Figure 8.12, the recognition data were consistent with Yonelinas's theoretical predictions.

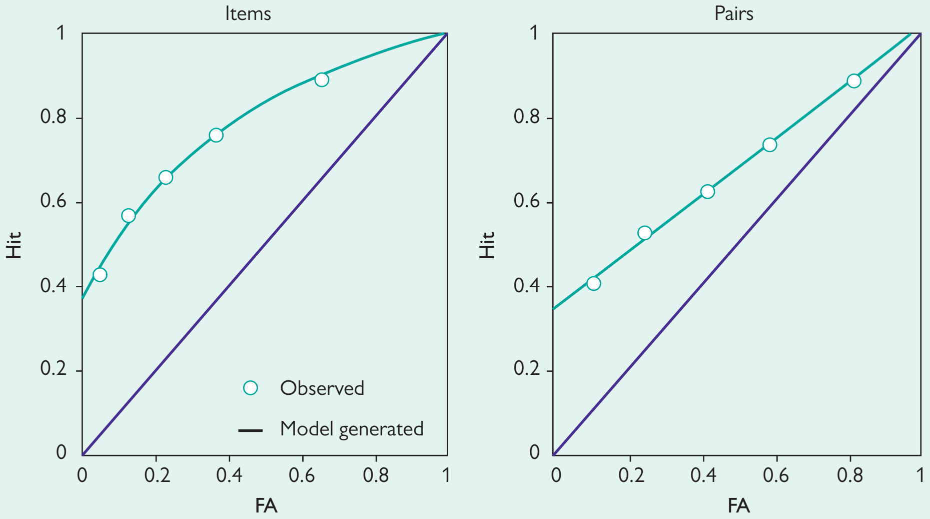


FIGURE 8.12 Recognition ROC curves for single items and item pairs. *Source: Yonelinas (1997)*

Free recall derives greatest benefit from relational processing, but little benefit from maintenance rehearsal (i.e. item-specific processing), which promotes integration. In contrast, familiarity-based recognition depends upon the degree of integration. Therefore, a high degree of item-specific processing (maintenance rehearsal) will benefit recognition to a greater extent than free recall. This illustrates that a full explanation of the effect on memory performance of different forms of encoding requires consideration of the relations between memory encoding, memory representation, and memory tests. This is the point made in Section 1 – whether interest is in encoding, storage, or retrieval, all stages of memory are involved when information is remembered.

4.4 Remembering and knowing

Yonelinas (2001) suggests that the distinction between remembering and knowing, first proposed by Tulving (1985) but explored in greater detail by Gardiner and associates (e.g. Gardiner, 1988; Gardiner and Java, 1990; Gardiner *et al.*, 1998, 2002; Gardiner and Richardson-Klavehn, 2000), reflects the different phenomenological experiences resulting from recollective and familiarity-based retrieval. Subjects are instructed to report a **remember experience** when they have a conscious recollection of the item appearing in the study. Subjects are instructed to report a **know experience** when they just know that the item appeared but have no conscious recollection of its occurrence (see Figure 8.13). Recollective retrieval is assumed to provide the qualitative information necessary for remembering, while familiarity-based retrieval provides a quantitative 'feeling' about the item, sometimes expressed as a sense of familiarity, but usually as a feeling of *knowing* the item was presented before.

Donaldson (1996; see also Dunn, 2004, 2008) argues that rather than reflecting qualitatively different memory retrieval processes and memory information, remember and know responses simply reflect decisions based on different response criteria applied

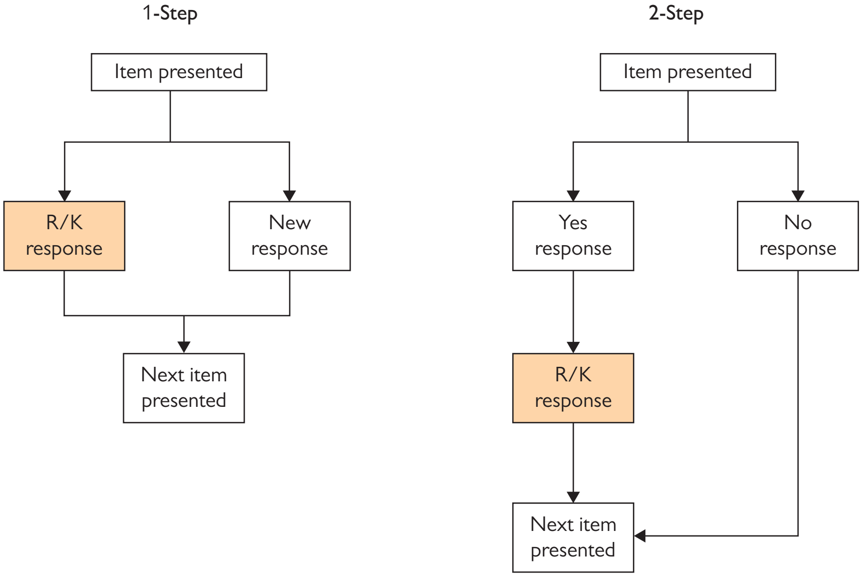


FIGURE 8.13 A schematic outline of the recognition one- and two-step remember (R)-know (K) paradigms. In one-step paradigms, subjects immediately provide R or K responses to recognized items, or they respond that the items are New. In two-step paradigms, subjects initially provide a standard yes–no recognition response and only when they respond yes are they required to provide a R or K response. (See Bruno and Rutherford, 2010, for the effects of including a Guess response option.)

to the strength of evidence axis: a response criterion high on the strength of evidence axis is used to identify remember items, while another response criterion lower on the strength of evidence axis is used to identify know items. If this account is accurate, then remember and know responses simply represent different levels on the single, strength of evidence dimension, and so only a single decision process operating on a single type of information is required. However, if remember and know responses tap into distinct recollection and familiarity-based retrieval processes, then two types of information and so two types of decision process will be required (e.g. Murdock, 2006; Reder *et al.*, 2000; Yonelinas, 2001).

Wixted (e.g. Wixted and Stretch, 2004; Wixted and Mickes, 2010) presents a modified version of Yonelinas's DPSD model. This assumes separate sources of recollective and familiarity-based information (also see Rotello *et al.*, 2004), but regards recollection as outputting a single continuous quantitative signal that is combined with the output from the familiarity-based retrieval process to provide a single continuous quantitative strength of evidence signal. All recognition and remember and know decisions are then based upon this signal. This is an active area of research and debate, that eventually should provide a much greater understanding of recognition memory (e.g. Mickes *et al.*, 2009; Parks and Yonelinas, 2007; Parks and Yonelinas, 2008; Slotnick, 2010; Wais *et al.*, 2008).

Although the remember-know task is a popular method for separating recollective and familiarity-based retrieval processes, its effective use requires some care (e.g. Geraci *et al.*, 2009; McCabe and Geraci, 2009; Mulligan *et al.*, 2010; Parks, 2007; Yonelinas and Jacoby, 1996). However, Yonelinas's DPSD account does not depend upon supporting evidence from remember-know task performance, or the analysis of ROC curves, as one of its strengths is the variety of evidence converging in its support (e.g. Vanna *et al.*, 2009). In particular, a consistent line of theorizing (e.g. Aggleton and Brown, 1999; Eichenbaum *et al.*, 2007) and a large and increasing body of neuroimaging research has linked distinct brain regions with recollective (remember) and familiarity-based (know) retrieval processing, and led Diana *et al.* (2007) and Ranganath (2010) to develop a 'binding of item and context' (BIC) model (see Figure 8.14).

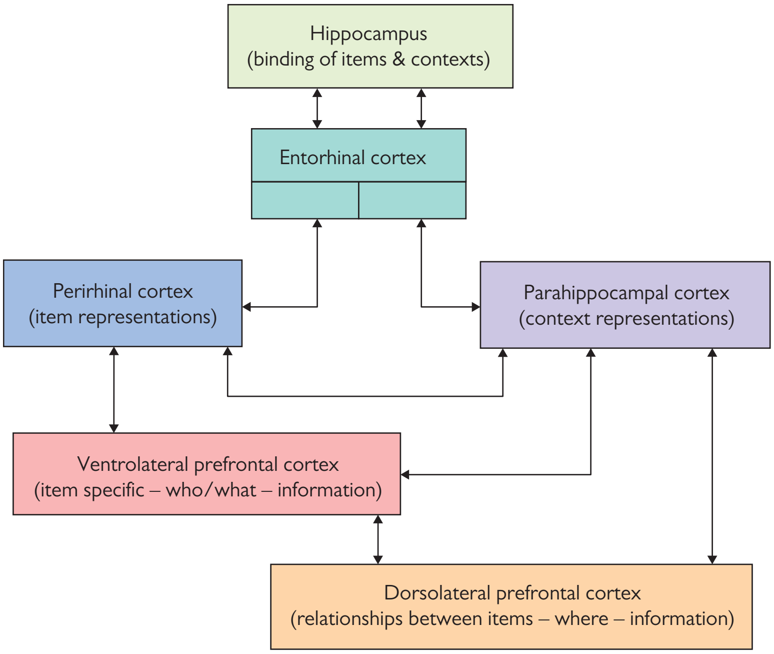

Hippocampus
(binding of items & contexts)
Entorhinal cortex
Perirhinal cortex
(item representations)
Parahippocampal cortex
(context representations)
Ventrolateral prefrontal cortex
(item specific – who/what – information)
Dorsolateral prefrontal cortex
(relationships between items – where – information)

FIGURE 8.14 The 'binding of item and context' (BIC) model. Arrows represent anatomical connections. *Source: adapted from Diana et al., 2007 and Ranganath, 2010*

According to the BIC model, item-specific (who and what) information and relational (where) information from the neocortex is sent via different anatomical routes to the MTL. The MTL perirhinal cortex processes the item-specific (who/what) information and the parahippocampal cortex processes the relational (where) information. Via the entorhinal cortex, these two structures output to the hippocampus, which binds together item and context information.

SUMMARY OF SECTION 4

- Retrieval involves finding, activating, and sometimes further processing pertinent memory representations.
- Both encoding specificity and TAP emphasize the encoding-retrieval relationship, with performance enhanced by increasing overlap of information (encoding specificity) or processing (TAP) at encoding and retrieval.
- The relative distinctiveness of a target item in memory and prior retrieval practice are important determinants of memory performance.
- Research is examining dual process accounts of recognition, which assume recollective and familiarity-based retrieval processes.
- Recognized (and recalled) items may provide a *remember* response (you consciously remember the item's presentation) or a know response (you know the item appeared but you have no conscious recollection of its presentation).
- Remember experiences and know experiences seem to be based on recollective and familiarity-based retrieval processes, respectively.
- Recollection decision processes assess qualitative information and employ a threshold process.
- Familiarity-based decision processes assess a continuous quantitative unidimensional signal and employ a modifiable SDT response criterion.
- A strength of the DPSD view of recollective and familiarity-based retrieval processing is the variety of evidence converging in its support.

5 IMPLICIT MEMORY

When free recall, cued recall, and recognition memory tests are employed, it is clear to subjects that they need to remember past events and so these are known as **explicit memory** tests. However, there are also **implicit memory** tests. According to Schacter (1987), implicit memory manifests when test performance is influenced by previous experience without the subject *intentionally* remembering past events or being *aware* of their relevance. In other words, implicit memory is supposed to be both unintentional and unconscious.

Several terms have been used to identify the type of memory test. These include: explicit versus implicit (Schacter and Tulving, 1994); direct versus indirect (Johnson and Hasher, 1987); and intentional versus incidental (Jacoby, 1984). As the intentionality aspect of implicit memory is less contested than the awareness aspect (e.g. Butler and Berry, 2001; Richardson-Klavehn, 2010), Jacoby's terminology will be applied, with explicit memory tests labelled as intentional tests and implicit memory tests labelled as incidental tests.

When "implicit memory" is used, it refers only to memory performance on incidental tests. Applying terminology that distinguishes between test types but avoids theoretical assumptions about the underlying memory organization is sensible when much of the research in this area addresses whether a processing or systems approach is most appropriate (see Section 5.3).

5.1 Perceptual and conceptual incidental tests

Roediger and McDermott (1993) list a variety of tests used to investigate *perceptual* and *conceptual* implicit memory. The word fragment task employed by Tulving *et al.* (1982) (see Box 8.9) is an example of a perceptual incidental test. Such tests usually require subjects to resolve displays that are perceptually impoverished (McDermott and Roediger, 1996) due to a short or an incomplete stimulus presentation. Examples include word-fragment completion (see Box 8.9) and anagram solution (e.g. Srinivas and Roediger, 1990), but also lexical decision – whether letter strings are real words (e.g. Duchek and Neely, 1989). **Conceptual incidental tests** require subjects to employ semantic knowledge to answer questions or respond to a cue (McDermott and Roediger, 1996). Examples include: word association (Shimamura and Squire, 1984); category exemplar generation (e.g. Srinivas and Roediger, 1990); and answering general knowledge questions (e.g. Blaxton, 1989). In all cases, implicit memory is demonstrated when test performance is better when test items were presented previously than when they were not.

BOX 8.9 RESEARCH STUDY Empirical evidence of implicit memory

Tulving *et al.* (1982) asked subjects to learn 96 words. Later they received a recognition test with 24 of the presented words (targets) and 24 similar distractors (i.e. words not presented before), and a word fragment completion test with 24 word fragments based on 24 presented words and 24 word fragments based on words not shown before (e.g. present the word fragment F_ O _ _ A _ L – subjects complete the fragment by placing O, T, B, and L in the empty letter slots to provide the word FOOTBALL). Seven days later, subjects received another recognition and word fragment completion test for the 48 words remaining. Critically, prior to both recognition tests subjects were instructed to respond 'yes' to words they had encountered at encoding. Prior to the word fragment tasks, subjects were instructed to respond with the first word that came to mind, with no reference made to the encoding session. The number of word fragments completed correctly when the corresponding word had been presented before was compared with the number of correct completions when no corresponding word had been presented before. A word repetition priming effect was observed – when words had been presented previously, more of the corresponding word fragments were completed, even though subjects had no appreciation that they were using memory.

Figure 8.15 shows the probability of a correct response as a function of test type (word fragment or recognition) after 1 hour and after 7 days. Unlike intentional recognition, there was no significant decline in incidental word fragment completion between the two test delays.

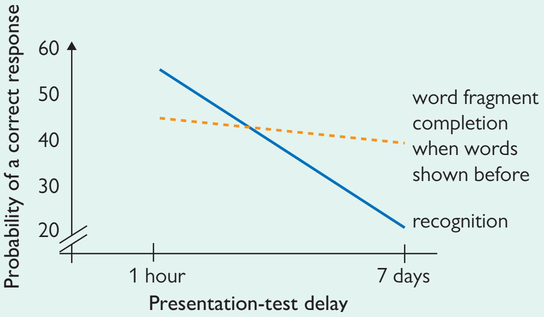


FIGURE 8.15 Probability of a correct response as a function of the presentation-test delay and type of test.

5.2 Intentionality and awareness

Implicit memory has been defined as memory retrieval without intention or awareness (Schacter, 1987). However, it is not clear to what extent the incidental tests described so far satisfy both aspects of this definition.

5.2.1 The retrieval intentionality criterion

One of the major advances in implicit memory research occurred with the development of the **retrieval intentionality criterion** (Schacter *et al.*, 1989) to address the issue of contamination from explicit memory (see Figure 8.16). The criterion distinguishes between intention to retrieve and awareness of retrieval. It also requires the same cues to be provided at retrieval for intentional and incidental tests, so the only difference between tests is the task instructions that manipulate retrieval intention. Consider the study by Tulving *et al.* (1982; see Box 8.9). A problem with this study is the test and retrieval instructions are confounded – subjects in the intentional retrieval conditions completed a recognition test, but subjects in the incidental retrieval conditions completed a word fragment test, and so it is impossible to claim that any differences observed are due only to the different retrieval instructions rather than the different cues provided at test.

Another requirement for the retrieval intentionality criterion is the identification of a variable that dissociates performance across the two tests. In the Tulving *et al.* (1982) study, a dissociation between intentional and incidental retrieval was observed: performance on the recognition test but not the word fragment completion test declined over time. If this dissociation across time had not been observed, then two interpretations would have been possible: i) there is no difference between intentional and incidental retrieval processes, or ii) there is no difference in performance between the intentional and incidental retrieval process tests because incidental test subjects actually employed intentional retrieval strategies. However, the observed dissociation between intentional and incidental test performance indicates that subjects in the word fragment condition did not use intentional retrieval strategies to complete the task. If intentional retrieval strategies had been used to complete both types of test, then there would be no basis for different memory performance across the two test

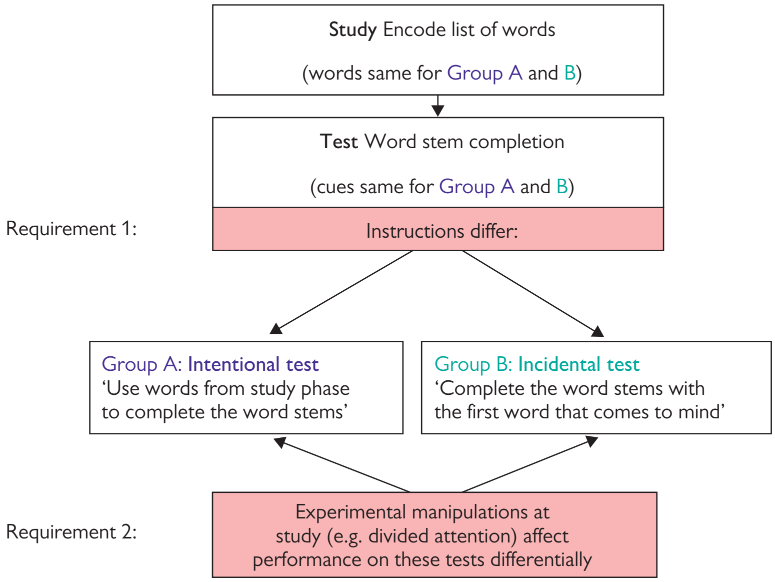

Study Encode list of words
(words same for Group A and B)
Test Word stem completion
(cues same for Group A and B)
Requirement 1:
Instructions differ:
Group A: Intentional test
'Use words from study phase to complete the word stems'
Group B: Incidental test
'Complete the word stems with the first word that comes to mind'
Requirement 2:
Experimental manipulations at study (e.g. divided attention) affect performance on these tests differentially

FIGURE 8.16 A schematic depiction of the retrieval intentionality criterion.

types – performance on the intentional recognition test and performance on the supposedly incidental, but actually intentional, word fragment test would decline similarly over time.

5.2.2 The issue of awareness in implicit memory

Functional dissociations have a key theoretical role in cognitive psychology (e.g. Dunn and Kirsner, 1988; Richardson-Klavehn and Bjork, 1988) and for implicit memory research in particular (see Section 5.3 and Richardson-Klavehn *et al.*, 2009). However, in implicit memory research, dissociations also play a vital methodological role in establishing the absence of explicit contamination in incidental memory tests (following the logic of the retrieval intentionality criterion just described). Many studies using incidental tests have employed LOP manipulations. Although LOP effects are observed frequently with intentional memory tests (e.g. Craik and Tulving, 1975; see Box 8.1), they are observed rarely with incidental tests (e.g. Graf and Mandler, 1984). Richardson-Klavehn and Gardiner (1996) investigated the effects of intentionality and awareness on retrieval using a LOP manipulation at encoding as an indicator of intentionality or explicit contamination. Awareness was measured item by item (for discussion of other measures see Butler and Berry, 2001 and Richardson-Klavehn *et al.*, 1996). Their data and a number of other studies (e.g. Richardson-Klavehn, 2010) demonstrate that conscious awareness of retrieval does not equate with intentional retrieval. A dissociation between awareness and intentionality has also been demonstrated in studies of conceptual implicit memory (e.g. Mulligan, 2002). Neuroimaging data also strongly indicate that awareness and retrieval intention are dissociable (e.g. Schott *et al.*, 2005).

5.3 Accounts of implicit memory

The distinction between incidental and intentional memory tasks provides a *description* of a person's psychological experience of memory use as different tasks are done (Schacter, 1987). However, it does not provide an *explanation* in terms of the cognitive processes involved. So far, most research has focused on transfer appropriate processing or memory systems accounts to explain these phenomena.

5.3.1 TAP account

Roediger and associates (e.g. Roediger *et al.*, 1989) have been the strongest advocates of Morris *et al.*'s (1977) TAP account as an explanation of performance differences on incidental and intentional memory tests. They argue that the match between the type of processing (perceptual or conceptual) engaged at encoding and when stimuli are tested is key, rather than differences between memory store retrieval characteristics. TAP predicts perceptual incidental tests will benefit most from encoding that engages similar perceptual processes. Likewise, performance on conceptual incidental tests will benefit more from encoding that engages similar conceptual processes. Indeed, perceptual manipulations at encoding have been shown to influence perceptual incidental memory test performance, but not conceptual intentional memory test performance (see Roediger and McDermott, 1993, for a review). Moreover, Blaxton (1989) demonstrated that generating words from a cue at study led to more conceptual priming than reading words, a finding consistent with the TAP account, as word generation is assumed to involve conceptual processing (cf. Mulligan, 2002; see Section 5.3.3). However, as the TAP account was examined, discrepant findings began to emerge. For example, Hunt *et al.* (1990) noted that orthographic distinctiveness (a perceptual factor) affected both (perceptual) incidental memory test performance and free recall.

Further evidence against the TAP account comes from studies of amnesic patients. Vaidya *et al.* (1995) found no difference between amnesics and controls in either a perceptual incidental memory test or a conceptual incidental memory test. However, amnesics' performance on intentional, perceptual, and conceptual memory tests was as poor as expected. Similar findings were reported by Cermak *et al.* (1995). The

TAP account of implicit memory attributes amnesics' ability to perform on conceptual incidental memory tests on a par with normal controls to their retained perceptual processing capability, while their poor memory performance is attributed to impaired conceptual processing. However, the ability of amnesics to perform on conceptual incidental memory tests on a par with normal controls contradicts this account. Moreover, the fact that amnesics exhibited their usual poor memory performance on intentional perceptual and conceptual memory tests suggests that the distinction between intentional and incidental memory tests is more critical than the distinction between perceptual and conceptual processing.

5.3.2 Memory systems accounts

The multiple memory systems and declarative and non-declarative memory systems perspectives have been applied to account for the differences between intentional and incidental memory test performance. In both cases, the differences are attributed to different memory systems being tested (see Figure 8.7 and Table 8.3). Squire (1992, 2007) simply attributes performance on intentional memory tests to the declarative memory system and performance on incidental memory tests to the non-declarative memory system. The multiple memory systems perspective attributes performance on perceptual incidental memory tests to the (unspecified) nature of the visual word form sub-system of the perceptual representation system, while picture priming is attributed to the (unspecified) nature of the structural description sub-system of the perceptual representation system (e.g. Schacter *et al.*, 2000). Within this perspective, Gabrieli (1999) attributes performance on conceptual incidental memory tests to yet another memory system – the conceptual representation system (cf. perceptual representation system; see Table 8.3).

In contrast to the memory systems perspectives, some researchers argue that the operation of a single memory system explains both incidental and intentional memory performance. Berry *et al.* (2008) constructed a SDT-based computational model that obtains a large number of dissociations by assuming the noise variance associated with incidental tests is larger than the noise variance associated with intentional retrieval tests. However, this model seems unable to account for the complex pattern of intentional and incidental memory test dissociations (Richardson-Klavehn, 2010; Richardson-Klavehn *et al.*, 1999; see also Section 5.3.3).

5.3.3 Components of processing account

To deal with the problems associated with the TAP interpretation of implicit memory, Roediger and associates modified their account and relabelled it *components of processing* (e.g. Roediger *et al.*, 1999). This view considers performance on different memory tests to involve different sets of processes. The sets of processes elicited by different memory tests may overlap, but some different process components will be elicited by any two tests that dissociate. However, the components of processing (COP) account has been criticized for being circular – it is assumed that priming occurs when there is appropriate transfer of processing, but the mark of appropriate transfer of processing is considered to be priming (see Section 2.1 for a similar criticism of LOP). Greater detail on the mechanisms operating in these circumstances is necessary to avoid this circularity, and indeed, this is one of the requirements McDermott and Roediger (1996) place on the components of processing account. Mulligan (2002) presented data suggesting that a full components of processing account required a distinction between item-specific and relational processing (see Section 2.2). He employed a number of different generation encoding tasks where subjects were presented with cues and then had to generate responses, and assumed the generated responses would be encoded as stimulus items. Generation tasks were employed at encoding because they have been shown to produce opposite effects on intentional and incidental memory tests (Jacoby, 1983). A semantic generation encoding task, where subjects generated a word based on a semantic cue (e.g. 'Lightning often occurs with t_____': the response item generated is expected to be 'thunder'), benefited both intentional retrieval and incidental conceptual retrieval – consistent with the COP account. However, perceptual generation encoding tasks (where subjects filled in the first two missing letters of each word) benefited intentional but not incidental conceptual retrieval, which is at odds with the COP account. Mulligan resolved this discrepancy by demonstrating that semantic generation tasks at encoding require both item-specific and relational processing, while non-semantic generation tasks involve mainly item-specific processing. At the retrieval stage, category-cued recall (an intentional conceptual test) benefits from both relational and item-specific processing, because the task requires the production of words that were on the study list *and* belong to the category, while the conceptual incidental test (i.e. category exemplar production) only benefits from relational processing, as the task does not require discrimination between studied and not studied exemplars. Consequently, a semantic generation encoding task affects both retrieval modes, as it involves both item-specific and relational processing, while a non-semantic generation task affects only category-cued recall due to the item-specific processing involved (and the lack of relational processing).

Richardson-Klavehn (2010) reviewed a number of studies presenting functional neuroimaging data and formulated an integrative theoretical framework of retrieval based on the components of processing account. The framework extends Moscovitch's (1992) account by postulating that all retrieval processes occur automatically and that different forms of memory arise from the way control processes (attention and working memory) interact with retrieval processes. One advantage of this account is its integration of diverse data patterns (behavioural and neurological) and existing theories of cognitive function from different research topics (e.g. working memory, attentional processes), making it both plausible and empirically testable. Nevertheless, more research is required to assess the framework assumptions.

SUMMARY OF SECTION 5

- Implicit retrieval was defined initially as unintentional and unconscious, but there is strong evidence that intentionality and awareness are dissociable.
- Perceptual incidental memory tests include: word fragment completion, word stem completion, word identification, anagram solution, and lexical decision.
- Conceptual incidental memory tests include: word association, category instance generation, and answering general knowledge questions.
- Amnesics exhibit normal memory performance on incidental memory tests.
- The multiple memory systems perspective and Squire's declarative and procedural memory systems attribute the differences between intentional and incidental memory test performance to different memory systems.
- The TAP account attributes the differences between intentional and incidental memory test performance to perceptual incidental tests benefiting most from perceptual encoding at presentation, while conceptual incidental tests and intentional memory tests benefit from conceptual encoding at presentation.
- The components of processing account can accommodate a large variety of different findings, but it suffers from circularity problems. However, these problems may be overcome by developing more specific models and predictions.

6 CONCLUSIONS

The systems and processes approaches are the two prevailing perspectives in cognitive psychology. Craik and Lockhart's LOP article stimulated a great deal of research on memory encoding processes and did much to establish the processing approach as a theoretical perspective in psychology. LOP research demonstrated the key role of encoding processing, but subsequent research by Mandler, Tulving, and Morris, Bransford, and Franks emphasized the role of retrieval processing and demonstrated that memory performance depends upon the interaction between encoding, representation, and retrieval. Since the original LOP article, there has been a substantial change in the conception of the nature of mental processing. Rather than being conceived only in terms of semantic degree or extent, processing is now conceived more broadly, with a range of processing modes available to subjects, some of which may engage "deep", but not semantic, processing. An appreciation that information can be encoded with reference to oneself is one example of this new conception of mental processing. This sort of encoding is extremely effective, probably due to the 'self' acting as a superordinate schema that influences encoding and retrieval operations (e.g. Van den Bos *et al.*, 2010).

Around the same time as Craik and Lockhart's levels of processing article, Tulving provided a description of separate semantic and episodic memory systems, but it was not until the 1980s that memory systems research began to exert substantial theoretical influence. This influence seems to have arisen as a consequence of a number of related factors, including a renaissance in connectionist research, cognitive neuropsychological investigation of abnormal memory as a consequence of brain damage, and developments in neuroimaging techniques. Memory systems research has benefited from the theoretical constraints imposed by neuropsychological findings and the theoretical criticism provided by cognitive psychologists. As research from a memory systems perspective (conducted primarily by neuropsychologists and neuroimaging researchers) and research from a memory processes perspective (conducted primarily by cognitive psychologists) progresses, the distinction between these two approaches appears to be diminishing. There is greater appreciation of the need to give account of the nature of processing within systems, and greater appreciation of the biological basis of systems (e.g. Squire, 2007) and how this can affect the nature of processing within systems.

A great deal of recent research activity has focused on retrieval operations. Yonelinas's DPSD model prompted considerable research activity across cognitive psychology, neuropsychology, and neuroimaging research. Neuropsychological and neuroimaging research has provided almost consistent support for Yonelinas's DPSD model, but opinion is still divided in cognitive psychology, where the nature of the recollective retrieval process is a major point of contention. Despite early consideration and rejection of DPSD models in favour of continuous detection models, only now are threshold detection models, their predictions, and the criteria by which they should be judged being examined with appropriate scrutiny (e.g. Algarabel and Pitarque, 2007; Broder and Schutz, 2009; Macho, 2004; Martin *et al.*, 2011; Meiser *et al.*, 2008). Although many matters concerning DPSD and SPSD models remain to be resolved, DPSD models seem to provide greater contact with other theoretical accounts and issues. Yonelinas's account of recollection – a threshold retrieval process that outputs qualitatively different types of information – offers a route to explain people's phenomenological experience of the remember component of recognition. In contrast, SPSD models retrieval processes output only a unidimensional quantitative (i.e. continuous) measure, and so are unable to provide convincing accounts of how qualitatively different memory information is retrieved, even when they are augmented by further formal memory model assumptions (e.g. SAM: Gillund and Shiffrin, 1984; Minerva 2: Hintzman, 1988). Links between free recall and recognition are often mentioned (e.g. Mandler, 1980) but, as SPSD models find free recall problematic, again such links seem more likely to be established by DPSD than SPSD models (e.g. Brainerd and Reyna, 2010).

Important theoretical and methodological progress in the study of implicit memory has been provided by the retrieval intentionality criterion. As this distinguishes between retrieval intention and retrieval awareness, it immediately revealed that previous implicit memory definitions were too simplistic. The theoretical perspective provided by the retrieval intentionality criterion also led to modified study designs, which provided more consistent data that was more easily interpretable within existing theoretical frameworks. The benefits of applying the retrieval intentionality criterion research can be seen in recent theoretical accounts, which incorporate a number of related aspects of memory, such as item-specific and relational processing, in the interpretation of the data.

An aim of this chapter was to present an overview of research in the psychology of memory that not only reflects theoretical influence and change, but also demonstrates the exciting advances in understanding that these new theoretical perspectives have provided. Memory research continues to be one of the most active research areas in psychology, where important theoretical and methodological developments are leading to a more accurate appreciation of memory operation.

FURTHER READING

Baddeley, A., Eysenck, M.W., and Anderson, M.C. (2009) *Memory*, Hove, Psychology Press. This is an up-to-date, scholarly, and accessible account of memory research.

Sedikides, C. and Spencer, S.J. (2007) *The Self*, New York, Psychology Press. A book that presents an up-to-date, scholarly, and accessible account of research on the self.

Richardson-Klavehn, A., Gardiner, J.M., and Java, R.I. (1996) 'Memory: task dissociations, process dissociations and dissociations of consciousness', in Underwood, G. (ed.) *Implicit Cognition*, Oxford, Oxford University Press. This chapter provides a good and clear discussion of intention and awareness issues, as well as various methods of measuring awareness.

Richardson-Klavehn, A. and Bjork, R.A. (2002) 'Memory: long term', in *Encyclopedia of Cognitive Science*, vol. 2 (pp.1096–105). The authors present a brief overview of LTM and also describe the concept of resonance.

REFERENCES

Aggleton, J.P. and Brown, M.W. (1999) 'Episodic memory, amnesia and the hippocampal–anterior thalamic axis', *Behavioural and Brain Sciences*, vol.22, pp.425–44.

Algarabel, S. and Pitarque, A. (2007) 'Context, remember-know recognition judgements, and ROC parameters', *Memory*, vol.15, pp.477–94.

Anderson, J.R. and Ross, B.H. (1980) 'Evidence against a semantic–episodic distinction', *Journal of Experimental Psychology: Human Learning and Memory*, vol.6, pp.441–65.

Atkinson, R.C. and Juola, J.F. (1973) 'Factors influencing speed and accuracy of word recognition', in Kornblum, S. (ed.) *Attention and Performance IV*, New York, Academic Press.

Atkinson, R.C. and Juola, J.F. (1974) 'Search and decision processes in recognition memory', in Kroutz, D.H., Atkinson R.C., and Suppes, P. (eds.) *Contemporary Developments in Mathematical Psychology*, San Francisco, Freeman.

Atkinson, R.C. and Shiffrin, R. (1968) 'Human memory: a proposed system and its control processes', in Spence, K. and Spence, J. (eds.) *The Psychology of Learning and Motivation*, vol.2, New York, Academic Press.

Baddeley, A.D. (1978) 'The trouble with levels: a re-examination of Craik and Lockhart's framework for memory research', *Psychological Review*, vol.85, pp.139–52.

Baddeley, A.D. (1984) 'Neuropsychological evidence and the semantic/episodic distinction [commentary]', *Behavioral and Brain Sciences*, vol.7, pp.238–9.

Baddeley, A.D. (2002) 'The concept of episodic memory' in Baddeley, A., Conway, M., and Aggleton, J. (eds.) *Episodic Memory: New Directions in Research*, Oxford, Oxford University Press.

Baddeley, A.D. and Warrington, E.K. (1970) 'Amnesia and the distinction between long- and short-term memory', *Journal of Verbal Learning and Verbal Behavior*, vol.9, pp.176–89.

Baddeley, A., Eysenck, M.W., and Anderson, M.C. (2009) *Memory*, Hove, Psychology Press.

Baddeley, A.D., Vargha-Khadem, F., and Mishkin, M. (2001) ‘Preserved recognition in a case of developmental amnesia: implications for the acquisition of semantic memory’, *Journal of Cognitive Neuroscience*, vol.13, pp.357–69.

Banks, W.P. (1970) ‘Signal detection theory and human memory’, *Psychological Bulletin*, vol.74, pp.81–99.

Bayley, P.J. and Squire, L.R. (2005) ‘Failure to acquire new semantic knowledge in patients with large medial temporal lobe lesions’, *Hippocampus*, vol.15, pp.273–80.

Bernard, F., Bullmore, E., Graham, K., Thompson, S., Hodges, J., and Fletcher, P. (2004) ‘The hippocampal region involved in successful recognition of both remote and recent famous faces’, *Neuroimage*, vol.22, pp.1704–14.

Berry, C.J., Shanks, D.R., and Henson, R.N.A. (2008) ‘A single system account of the relationship between priming, recognition and fluency’, *Journal of Experimental Psychology: Human Learning and Memory*, vol.34, pp.97–111.

Blaxton, T.A. (1989) ‘Investigating dissociations among memory measures: support for a transfer-appropriate processing framework’, *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.15, pp.657–68.

Bower, G.H. (1972) ‘Mental imagery and associative learning’, in Gregg, L.W. (ed.) *Cognition in Learning and Memory*, New York, Wiley.

Brainerd, C.J. and Reyna, V.F. (2010) ‘Recollective and non-recollective recall’, *Journal of Memory and Language*, vol.63, pp.425–45.

Brandt, K.R., Gardiner, J.M., Vargha-Khadem, F., Baddeley, A.D., and Mishkin, M. (2006) ‘Using semantic memory to boost “episodic” recall in a case of developmental amnesia’, *Neuroreport*, vol.17, pp.1057–60.

Bransford, J.D., Franks, J.J., Morris, C.D., and Stein, B.S. (1979) ‘Some general constraints on learning and memory research’, in Cermak, L.S. and Craik, F.I.M. (eds.) *Levels of Processing in Human Memory*, Hillsdale, NJ, LEA.

Bröder, A. and Schütz, J. (2009) ‘Recognition ROCs are curvilinear – or are they? On premature arguments against the two high threshold models of recognition’, *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.35, pp.587–606.

Bruno, D. and Rutherford, A. (2010) ‘How many response options? A study of remember–know testing procedures’, *Acta Psychologica*, vol.134, pp.125–9.

Bruno, D., Higham, P.A., and Perfect, T.J. (2009) ‘Global subjective memorability and the strength-based mirror effect in recognition memory’, *Memory and Cognition*, vol.37, pp.807–18.

Burianova, H. and Grady, C.L. (2007) ‘Common and unique neural activations in autobiographical, episodic, and semantic retrieval’, *Journal of Cognitive Neuroscience*, vol.19, pp.1520–34.

Butler, L.T. and Berry, D.C. (2001) ‘Implicit memory: intention and awareness revisited’, *Trends in Cognitive Sciences*, vol.5, pp.192–7.

Cabeza, R. and Nyberg, L. (2000) ‘Imaging cognition ii: an empirical review of 275 PET and fMRI studies’, *Journal of Cognitive Neuroscience*, vol.12, pp.1–47.

Carr, V., Rissman, J., and Wagner, A. (2010) ‘Imaging the human medial temporal lobe with high-resolution fMRI’, *Neuron*, vol.65, pp.298–308.

Cermak, L.S. and O’Connor, M. (1983) ‘The anterograde and retrograde retrieval ability of a patient with amnesia due to encephalitis’, *Neuropsychologia*, vol.21, pp.213–34.

Cermak, L.S., Verfaelie, M., and Chase, K.A. (1995) ‘Implicit and explicit memory in amnesia: an analysis of data-driven and conceptually driven processes’, *Neuropsychology*, vol.9, pp.281–90.

Cohen, N.J. and Squire, L.S. (1980) ‘Preserved learning and retention of pattern-analysing skill in amnesia: dissociation of knowing how and knowing that’, *Science*, vol.210, pp.207–10.

Craik, F.I.M. (1977) ‘“A level of analysis” view of memory’, in Pliner, P., Krames, L., and Allaway, T.M. (eds.) *Communication and Affect: Vol 2. Language and Thought*, New York, Academic Press.

Craik, F.I.M. (2002) ‘Levels of processing: past, present . . . and future?’ *Memory*, vol.10, pp.305–18.

Craik, F.I.M. and Lockhart, R.S. (1972) ‘Levels of processing: a framework for memory research’, *Journal of Verbal Learning and Verbal Behavior*, vol.11, pp.671–84.

Craik, F.I.M. and Tulving, E. (1975) ‘Depth of processing and the retention of words in episodic memory’, *Journal of Experimental Psychology: General*, vol.104, pp.268–94.

Crowder, R. (1993) ‘Systems and principles in memory theory: another critique of pure memory’, in Collins, A.F., Gathercole, S.E., Conway, M.A., and Morris, P.E. (eds.) *Theories of Memory*, Hove, LEA.

Diana, R.A., Yonelinas, A.P., and Ranganath, C. (2007) ‘Imaging recollection and familiarity in the medial temporal lobe: a three component model’, *Trends in Cognitive Sciences*, vol.11, pp.379–86.

Donaldson, W. (1996) ‘The role of decision processes in remembering and knowing’, *Memory and Cognition*, vol.24, pp.523–33.

Duchek, J.M. and Neely, J.H. (1989) ‘A dissociative word-frequency X levels-of-processing interaction in episodic recognition and lexical decision tasks’, *Memory and Cognition*, vol.17, pp.148–62.

Dunn, J.C. (2004) ‘Remember-Know: a matter of confidence’, *Psychological Review*, vol.111, pp.524–42.

Dunn, J.C. (2008) ‘The dimensionality of the remember-know task: a state-trace analysis’, *Psychological Review*, vol.115, pp.426–46.

Dunn, J.C. and Kirsner, K. (1988) 'Discovering functionally independent mental processes: the principle of reversed association', *Psychological Review*, vol.95, pp.91–101.

Eich, J.M. (1985) 'Levels of processing, encoding specificity, elaboration and CHARM', *Psychological Review*, vol.92, pp.1–38.

Eichenbaum, H., Yonelinas, A.P., and Ranganath, C. (2007) 'The medial temporal lobe and recognition memory', *Annual Review of Neuroscience*, vol.30, pp.123–52.

Eysenck, M.W. and Eysenck, M.C. (1980) 'Effects of processing depth, distinctiveness and word frequency on retention', *British Journal of Psychology*, vol.71, pp.263–74.

Eysenck, M.W. and Keane, M.T. (2010) *Cognitive Psychology: A Student's Handbook*, Hove, Psychology Press.

Foster, J.K. and Jelic, M. (1999) *Memory: Systems, Process or Function?*, Oxford, Oxford University Press.

Gabrieli, J.D.E. (1999) 'The architecture of human memory', in Foster, J.K. and Jelic, M. (eds.) *Memory: Systems, Process or Function?*, Oxford, Oxford University Press.

Gabrieli, J.D.E., Cohen, N.J., and Corkin, S. (1988) 'The impaired learning of semantic knowledge following bilateral medial-temporal lobe resection', *Brain*, vol.7, pp.157–77.

Gardiner, J.M. (1988) 'Functional aspects of recollective experience', *Memory and Cognition*, vol.16, pp.309–13.

Gardiner, J.M. and Java, R.I. (1990) 'Recollective experience in word and nonword recognition', *Memory and Cognition*, vol.18, pp.23–30.

Gardiner, J.M. and Richardson-Klavehn, A. (2000) 'Remembering and knowing', in Tulving, E. and Craik, F.I.M. (eds.) *The Oxford Handbook of Memory*, Oxford, Oxford University Press.

Gardiner, J.M., Ramponi, C., and Richardson-Klavehn, A. (1998) 'Experiences of remembering, knowing and guessing', *Consciousness and Cognition*, vol.7, pp.1–26.

Gardiner, J.M., Ramponi, C., and Richardson-Klavehn, A. (2002) 'Recognition memory and decision processes: a meta-analysis of remember, know and guess responses', *Memory*, vol.10, pp.83–98.

Geraci, L., McCabe, D.P., and Guillory, J.J. (2009) 'On interpreting the relationship between remember–know judgments and confidence: the role of instructions', *Consciousness and Cognition*, vol.18, pp.701–09.

Gillund, G. and Shiffrin, R.M. (1984) 'A retrieval model for both recognition and recall', *Psychological Review*, vol.91, pp.1–67.

Glenberg, A.M., Smith, S.M., and Green, C. (1977) 'Type 1 rehearsal: maintenance and more', *Journal of Verbal Learning and Verbal Behavior*, vol.16, pp.339–52.

Graf, P. and Mandler, G. (1984) 'Activation makes words more accessible, but not necessarily more retrievable', *Journal of Verbal Learning and Verbal Behavior*, vol.23, pp.553–68.

Green, D.M. and Swets, J.A. (1966) *Signal Detection Theory and Psychophysics*, New York, Wiley.

Greenberg, D.L. and Verfaellie, M. (2010) 'Interdependence of episodic and semantic memory: evidence from neuropsychology', *Journal of the International Neuropsychological Society*, vol.16, pp.1–6.

Grossman, M. (1987) 'Lexical acquisition in alcoholic Korsakoff psychosis', *Cortex*, vol.23, pp.631–44.

Henke, K., Mondadori, C., Treyer, V., Nitsch, R., Buck, A., and Hock, C. (2003) 'Nonconscious formation and reactivation of semantic associations by way of the medial temporal lobe', *Neuropsychologia*, vol.41, pp.863–76.

Hintzman, D.L. (1986) 'Schema abstraction in a multiple-trace memory model', *Psychological Review*, vol.93, pp.411–28.

Hintzman, D.L. (1988) 'Judgements of frequency and recognition in a multiple-trace memory model', *Psychological Review*, vol.95, pp.528–51.

Hoscheidt, S.M., Nadel, L., Payne, J., and Ryan, L. (2010) 'Hippocampal activation during retrieval of spatial context from episodic and semantic memory', *Behavioural Brain Research*, vol.212, pp.121–32.

Humphreys, M.S., Bain, J.D., and Pike, R. (1989) 'Different ways to cue a coherent memory system: a theory for episodic, semantic and procedural tasks', *Psychological Review*, vol.96, pp.208–33.

Hunt, R.R. and Einstein, G.O. (1981) 'Relational and item-specific information in memory', *Journal of Verbal Learning and Verbal Behavior*, vol.20, pp.497–514.

Hunt, R.R. and McDaniel, M.A. (1993) 'The enigma of organization and distinctiveness', *Journal of Memory and Language*, vol.32, pp.421–45.

Hunt, R.R., Humphrey, N., and Toth, J.P. (1990) 'Perceptual identification, fragment completion and free recall: concepts and data', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.16, pp.282–90.

Jacoby, L.L. (1983) 'Remembering the data: analyzing interactive processes in reading', *Journal of Verbal Learning and Verbal Behavior*, vol.22, pp.485–508.

Jacoby, L.L. (1984) 'Incidental versus intentional retrieval: remembering and awareness as separate issues', in Squire, L.R. and Butters, N. (eds.) *Neuropsychology of Memory*, New York, Guildford Press.

Jacoby, L.L. and Dallas, M. (1981) 'On the relationship between autobiographical memory and perceptual learning', *Journal of Experimental Psychology: General*, vol.3, pp.306–40.

Johnson, M.K. and Hasher, L. (1987) 'Human learning and memory', *Annual Review of Psychology*, vol.38, pp.631–68.

Koffka, K. (1935) *Principles of Gestalt Psychology*, New York, Harcourt Brace.

Leveroni, C., Seidenberg, M., Mayer, A., Mead, L., Binder, J., and Rao, S. (2000) 'Neural systems underlying the recognition of familiar and newly learned faces', *Journal of Neuroscience*, vol.20, pp.878–86.

Lockhart, R.S. and Murdock, B.B. (1970) 'Memory and the theory of signal detection', *Psychological Bulletin*, vol.74, pp.100–09.

Lockhart, R.S., Craik, F.I.M., and Jacoby, L.L. (1976) 'Depth of processing, recognition and recall', in Brown, J. (ed.) *Recall and Recognition*, London, Wiley.

Macho, S. (2004) 'Modeling associative recognition: a comparison of two-high-threshold, two-high-threshold signal detection, and mixture distribution models', *Journal of Experimental Psychology: Learning, Memory & Cognition*, vol.30, pp.83–97.

Macmillan, N.A. and Creelman, C.D. (2005) *Detection Theory: A User's Guide*, Mahwah, NJ, Erlbaum.

Maguire, E. and Mummery, C. (1999) 'Differential modulation of a common memory retrieval network revealed by positron emission tomography', *Hippocampus*, vol.9, pp.54–61.

Mandler, G. (1979) 'Organization and repetition: organizational principles with special reference to rote learning', in Nilsson, L-G. (ed.) *Perspectives on Memory Research*, Hillsdale, NJ, LEA.

Mandler, G. (1980) 'Recognizing: the judgement of previous occurrence', *Psychological Review*, vol.87, pp.252–71.

Mandler, G. (2008) 'Familiarity breeds attempts: a critical review of dual-process theories of recognition', *Perspectives on Psychological Science*, vol.3, pp.390–9.

Markopoulos, G., Rutherford, A., Cairns, C., and Green, J. (2010) 'Encoding instructions and stimulus presentation in local environmental context-dependent memory studies', *Memory*, vol.18, pp.610–24.

Martin, C.D., Baudouin, J.Y., Franck, N., Guillaume, F., Guillem, F., Huron, C., and Tiberghien, G. (2011) 'Comparison of RK and confidence judgement ROCs in recognition memory', *Journal of Cognitive Psychology*, vol.23, pp.171–84.

Matzen, L.E. and Benjamin, A.S. (2009) 'Remembering words not presented in sentences: how study context changes patterns of false memories', *Memory and Cognition*, vol.37, pp.52–64.

McCabe, D.P. and Geraci, L. (2009) 'The influence of instructions and terminology on the accuracy of remember–know judgments', *Consciousness and Cognition*, vol.18, pp.401–13.

McClelland, J.L. and Chappel, M. (1998) 'Familiarity breeds differentiation: a subjective likelihood approach to the effects of experience in recognition memory', *Psychological Review*, vol.105, pp.724–60.

McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995) 'Why there are complimentary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory', *Psychological Review*, vol.102, pp.419–57.

McDermott, K.B. and Roediger, H.L. (1996) 'Exact and conceptual repetition dissociate conceptual memory tests: problems for transfer appropriate processing theory', *Canadian Journal of Experimental Psychology*, vol.50, pp.57–71.

McKoon, G., Ratcliff, R., and Dell, G.S. (1986) 'A critical examination of the semantic/episodic distinction', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.12, pp.295–306.

Meiser, T., Sattler, C., and Weisser, K. (2008) 'Binding of multidimensional context information as a distinctive characteristic of Remember judgments', *Journal of Experimental Psychology: Learning Memory & Cognition*, vol.34, pp.32–49.

Mickes, L., Johnson, E.M., and Wixted, J. (2010) 'Continuous recollection versus unitized familiarity in associative recognition', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.36, pp.843–63.

Mickes, L., Wais, P.E., and Wixted, J. (2009) 'Recollection is a continuous process: implications for dual-process theories of recognition memory', *Psychological Science*, vol.20, pp.509–15.

Mickes, L., Wixted, J., and Wais, P.E. (2007) 'A direct test of the unequal-variance signal-detection model of recognition memory', *Psychonomic Bulletin & Review*, vol.14, pp.858–65.

Milner, B. (1966) 'Amnesia following operation on the temporal lobes', in Whitty, C.W.M. and Zangwill, O.L. (eds.) *Amnesia*, London, Butterworths.

Morris, C.D., Bransford, J.D., and Franks, J.J. (1977) 'Levels of processing versus transfer appropriate processing', *Journal of Verbal Learning and Verbal Behavior*, vol.16, pp.519–33.

Morton, J., Hammersley, R.H., and Bekerian, D.A. (1985) 'Headed records: a model for memory and its failures', *Cognition*, vol.20, pp.1–23.

Moscovitch, M. (1992) 'Memory and working-with-memory – a component process model based on modules and central systems', *Journal of Cognitive Neuroscience*, vol.4, pp.257–67.

Mulligan, N.W. (2002) 'The effects of generation on conceptual implicit memory', *Journal of Memory and Language*, vol.47, pp.327–42.

Mulligan, N.W., Besken, M., and Peterson, D. (2010) 'Remember-Know and source memory instructions can qualitatively change old-new recognition accuracy: the modality-match effect in recognition memory', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.36, pp.558–66.

Murdock, B.B., Jr. (1993) 'Todam 2 – a model for the storage and retrieval of item, associative, and serial-order information', *Psychological Review*, vol.100, pp.183–203.

Murdock, B. (2006) 'Decision-making models of remember-know judgments: comment on Rotello, Macmillan and Reeder (2004)', *Psychological Review*, vol.113, pp.648–55.

Nairne, J.S. (2002) 'The myth of the encoding-retrieval match', *Memory*, vol.10, pp.389–95.

Neely, J.H. (1989) 'Experimental dissociations and the semantic/episodic memory distinction', in Roediger, H.L. and Craik, F.I.M. (eds.) *Varieties of Memory and Consciousness: Essays in Honor of Endel Tulving*, Hillsdale, NJ, LEA.

Norman, D.A. (1982) *Learning and Memory*, San Francisco, Freeman.

Norman, D.A. and Bobrow, D.G. (1979) 'Descriptions: an intermediate stage in memory retrieval', *Cognitive Psychology*, vol.11, pp.107–23.

O'Kane, G., Kensinger, E.A., and Corkin, S. (2004) 'Evidence for semantic learning in profound amnesia: an investigation with patient H.M.', *Hippocampus*, vol.14, pp.417–25.

Parkin, A. (1993) *Human Memory*, Oxford, Blackwell.

Parks, C.M. (2007) 'The role of noncriterial recollection in estimating recollection and familiarity', *Journal of Memory and Language*, vol.57, pp.81–100.

Parks, C.M. and Yonelinas, A.P. (2007) 'Moving beyond pure signal-detection models: comment on Wixted', *Psychological Review*, vol.114, pp.188–202.

Parks, C.M. and Yonelinas, A.P. (2008) 'Theories of recognition memory', in Byrne, J.H. (chief ed.) *Learning and Memory: A Comprehensive Reference*, vol.2, Roediger, H. (ed.) *Cognitive Psychology of Learning*, Oxford, UK, Elsevier.

Parks, T.E. (1966) 'Signal-detectability theory of recognition-memory performance', *Psychological Review*, vol.73, pp.44–58.

Patterson, K., Nestor, P.J., and Rogers, T.T. (2007) 'Where do you know what you know? The representation of semantic knowledge in the human brain', *Nature Reviews Neuroscience*, vol.8, pp.976–9.

Prince, S.E., Tsukiura, T., and Cabeza, R. (2007) 'Distinguishing the neural correlates of episodic memory encoding and semantic memory retrieval', *Psychological Science*, vol.18, pp.144–51.

Rajah, M.N. and McIntosh, A.R. (2005) 'Overlap in the functional neural systems involved in semantic and episodic memory retrieval', *Journal of Cognitive Neuroscience*, vol.17, pp.470–82.

Ranganath, C. (2010) 'Binding items and contexts: the cognitive neuroscience of episodic memory', *Current Directions in Psychological Science*, vol.19, pp.131–37.

Ratcliffe, R., Sheu, C-F., and Gronlund, S.D. (1992) 'Testing global memory models using ROC curves', *Psychological Review*, vol.99, pp.518–35.

Reder, L.M., Nhouyvanisvong, A., Schunn, C.D., Ayers, M.S., Angstadt, P., and Hiraki, K. (2000) 'A mechanistic account of the mirror effect for word frequency: a computational model of remember-know judgments in a continuous recognition paradigm', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.26, pp.294–320.

Richardson-Klavehn, A. (2010) 'Priming, automatic recollection, and control of retrieval: toward an integrative retrieval architecture', in Mace, J.H. (ed.) *The Act of Remembering: Toward an Understanding of How We Recall the Past*, New York, Wiley-Blackwell.

Richardson-Klavehn, A. and Bjork, R.A. (1988) 'Measures of memory', *Annual Review of Psychology*, vol.39, pp.475–543.

Richardson-Klavehn, A. and Gardiner, J.M. (1996) 'Cross-modality priming in stem completion reflects conscious memory but not voluntary memory', *Psychonomic Bulletin and Review*, vol.3, pp.238–44.

Richardson-Klavehn, A., Bergström, Z.M., Magno, E., Markopoulos, G., Sweeney-Reed, C.M., and Wimber, M. (2009) 'On the intimate relationship between neurobiology and function in the theoretical analysis of human learning and memory', in Rösler, F., Ranganath, C., Röder, B., and Kluwe, R.H. (eds.) *Neuroimaging of Human Memory: Linking Cognitive Processes to Neural Systems*, New York, Oxford University Press.

Richardson-Klavehn, A., Clarke, A.J.B., and Gardiner, J.M. (1999) 'Conjoint dissociations reveal involuntary "perceptual" priming from generating at study', *Consciousness and Cognition*, vol.8, pp.271–84.

Richardson-Klavehn, A., Gardiner, J.M., and Java, R.I. (1996) 'Memory: task dissociations, process dissociations and dissociations of consciousness', in Underwood, G. (ed). *Implicit Cognition*, Oxford, Oxford University Press.

Roediger, H.L. and McDermott, K.B. (1993) 'Implicit memory in normal human subjects', in Boller, F. and Grafman, J. (eds.) *Handbook of Neuropsychology*, vol.8, Amsterdam, Elsevier.

Roediger, H.L., Buckner, R.L., and McDermott, K.B. (1999) 'Components of processing', in Foster, J.K. and Jelic, M. (eds.) *Memory: Systems, Process or Function?*, Oxford, Oxford University Press.

Roediger, H.L., Rajaram, S., and Srinivas, K. (1990) 'Specifying criteria for postulating memory systems', in Diamond, A. (ed.) *The Development and Neural Bases of Higher Cognitive Functions*, New York, New York Academy of Sciences.

Roediger, H.L., Weldon, M.S., and Challis, B.H. (1989) 'Explaining dissociations between implicit and explicit measures of retention: a processing account', in Roediger, H.L. and Craik, F.I.M. (eds.) *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving*, Hillsdale, NJ, LEA.

Rolls, E.T. and Kesner, R.P. (2006) 'A computational theory of hippocampal function and empirical tests of the theory', *Progress in Neurobiology*, vol.79, pp.1–48.

Rotello, C.M., Macmillan, N.A., and Reeder, J.A (2004) 'Sum-difference theory of remembering and knowing: a two-dimensional signal-detection model', *Psychological Review*, vol.111, pp.588–616.

Rundus, D. (1977) 'Maintenance rehearsal and single level processing', *Journal of Verbal Learning and Verbal Behaviour*, vol.16, pp.665–81.

Rutherford, A. (2000) 'The ability of familiarity, disruption and the relative strength of non-EC memory cues to explain unreliable environmental context-dependent memory effects in free-recall', *Memory and Cognition*, vol.28, pp.1419–28.

Ryan, L., Hoscheidt, S., and Nadel, L. (2008) 'Perspectives on semantic and episodic memory retrieval', in Dere, E., Easton, A., Nadel, L., and Huston, J.P. (eds.) *Handbook of Behavioural Neuroscience Series*, vol.18, *Handbook of Episodic Memory*, Netherlands, Elsevier Sciences.

Ryan, L., Lin, C-Y., Ketcham, K., and Nadel, L. (2010) 'The role of medial temporal lobe in retrieving spatial and nonspatial relations from episodic and semantic memory', *Hippocampus*, vol.139, pp.11–18.

Ryle, G. (1949) *The Concept of Mind*, London, Hutchinson.

Schacter, D.L. (1987) 'Implicit memory: history and current status', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.13, pp.501–18.

Schacter, D.L. and Tulving, E. (1994) 'What are the memory systems of 1994?' in Schacter, D.L. and Tulving, E. (eds.) *Memory Systems*, Cambridge, MA, MIT Press.

Schacter, D.L., Bowers, J., and Booker, J. (1989) 'Intention, awareness, and implicit memory: the retrieval intentionality criterion', in Lewandowsky, S., Dunn, J.C., and Kirsner, K. (eds.) *Implicit Memory*, Hillsdale, NJ, Erlbaum.

Schacter, D.L., Wagner, A.D., and Buckner, R.L. (2000) 'Memory systems of 1999′, in Tulving, E. and Craik, F.I.M. (eds.) *The Oxford Handbook of Memory*, Oxford, Oxford University Press.

Schott, B.H., Henson, R.N., Richardson-Klavehn, A., Becker, C., Thoma, V., Heinze, H-J., and Düzel, E. (2005) 'Redefining implicit and explicit memory: the functional neuroanatomy of priming, remembering, and control of retrieval', *Proceedings of the National Academy of Sciences USA*, vol.102, pp.1257–62.

Scoville, W.B. and Milner, B. (1957) 'Loss of recent memory after bilateral hippocampal lesions', *Journal of Neurology, Neurosurgery and Psychiatry*, vol.20, pp.11–21.

Shanks, D.R. (1997) 'Dissociating long-term memory systems: comment on Nyberg and Tulving (1996)', *European Journal of Cognitive Psychology*, vol.9, pp.111–20.

Shiffrin, R.M. and Steyvers, M. (1997) 'A model for recognition memory: REM-retrieving effectively from memory', *Psychological Bulletin and Review*, vol.4, pp.145–66.

Shimamura, A.P. and Squire, L.R. (1984) 'Paired associate learning and priming effects in amnesia: a neuropsychological approach', *Journal of Experimental Psychology: General*, vol.113, pp.556–70.

Slotnick, S.D. (2010) 'Remember source memory ROCs indicate recollection is a continuous process', *Memory*, vol.18, pp.27–39.

Spiers, H.J., Maguire, E.A., and Burgess, N. (2001) 'Hippocampal amnesia', *Neurocase*, vol.7, pp.357–82.

Squire, L.R. (1987) *Memory and Brain*, New York, Oxford University Press.

Squire, L.R. (1992) 'Declarative and nondeclarative memory: multiple brain systems supporting learning and memory', *Journal of Cognitive Neuroscience*, vol.4, pp.232–43.

Squire, L.R. (2007) 'Memory systems: a biological concept', in Roediger, H.L., Dudai, Y., and Fitzpatrick, S.M. (eds.) *Science of Memory: Concepts*, Oxford, Oxford University Press.

Squire, L., Stark, C., and Clark, R. (2004) 'The medial temporal lobe', *Annual Review of Neuropsychology*, vol.27, pp.279–306.

Srinivas, K. and Roediger, H.L. (1990) 'Classifying implicit memory tests: category association and anagram solution', *Journal of Memory and Language*, vol.29, pp.389–412.

Surprenant, A.M. and Neath, I. (2009) *Principles of Memory*, Hove, Psychology Press.

Swets, J.A. (1959) 'Indices of signal detectability obtained with various psychophysical procedures', *Journal of the Acoustical Society of America*, vol.31, pp.511–13.

Symons, C.S. and Johnson, B.T. (1997) 'The self reference effect in memory: a meta-analysis', *Psychological Bulletin*, vol.121, pp.371–94.

Thompson-Schill, S.L., Kan, I.P., and Oliver, R.T. (2006) 'Functional neuroimaging of semantic memory', in Cabeza, R. and Kingstone, A. (eds.) *Handbook of Functional Neuroimaging of Cognition*, 2nd edition, Cambridge, MA, MIT Press.

Toth, J.P. and Hunt, R.R. (1999) 'Not one versus many, but zero versus any; structure and function in the context of the multiple memory systems debate', in Foster, J.K. and Jelic, M. (eds.) *Memory: Systems, Process or Function?*, Oxford, Oxford University Press.

Tulving, E. (1972) 'Episodic and semantic memory', in Tulving, E. and Donaldson, W. (eds.) *Organization of Memory*, New York, Academic Press.

Tulving, E. (1979) 'Relation between encoding specificity and levels of processing', in Cermak, L.S. and Craik, F.I.M. (eds.) *Levels of Processing in Human Memory*, Hillsdale, NJ, LEA.

Tulving, E. (1983) *Elements of Episodic Memory*, Oxford, Clarendon Press.

Tulving, E. (1984) 'Précis of elements of episodic memory', *Behavioral and Brain Sciences*, vol.7, pp.223–68.

Tulving, E. (1985) 'Memory and consciousness', *Canadian Psychology*, vol.26, pp.1–12.

Tulving, E. and Osler, S. (1968) 'Effectiveness of retrieval cues in memory for words', *Journal of Experimental Psychology*, vol.77, pp.593–601.

Tulving, E., Schacter, D.L., and Stark, H.A. (1982) 'Priming effects in word-fragment completion are independent of recognition memory', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.8, pp.336–42.

Vaidya, C.J., Gabrieli, J.D.E, Keane, M.M., and Monti, L.A. (1995) 'Perceptual and conceptual memory processes in global amnesia', *Neuropsychology*, vol.10, pp.529–37.

Van den Bos, M., Cunningham, S.J., Conway, M.A., and Turk, D.J. (2010) 'Mine to remember: the impact of ownership on recollective experience', *The Quarterly Journal of Experimental Psychology*, vol.63, pp.1065–71.

Vanna, S.D., Tsivilis, D., Denby, C.E., Quamme, J.R., Yonelinas, A.P., Aggleton, J.P., Montaldi, D., and Mayes, A.R. (2009) 'Impaired recollection but spared familiarity in patients with extended hippocampal system damage revealed by 3 convergent methods', *Proceedings of the National Academy of Sciences*, vol.106, pp.5442–7.

Vargha-Khadem, F., Gadian, G.D., Watkins, K.E., Connelly, A., and van Paesschen, W. (1997) 'Differential effects of early hippocampal pathology on episodic and semantic memory', *Science*, vol.277, pp.376–80.

Verfaillie, M., Koseff, P., and Alexander, M.P. (2000) 'Acquisition of novel semantic information in amnesia: effects of lesion location', *Neuropsychologia*, vol.38, pp.484–92.

Wais, P.E., Mickes, L., and Wixted, J.T. (2008) 'Remember/know judgments probe degrees of recollection', *Journal of Cognitive Neuroscience*; vol.20, pp.400–5.

Wickens, T.D. (2002) *Elementary Signal Detection Theory*, Oxford, Oxford University Press.

Wixted, J.T. and Mickes, L. (2010) 'A continuous dual-process model of remember/know judgments', *Psychological Review*, vol.117, pp.1025–54.

Wixted, J.T. and Stretch, V. (2004) 'In defense of the signal detection interpretation of remember/know judgments', *Psychonomic Bulletin and Review*, vol.11, pp.616–41.

Yonelinas, A.P. (1994) 'Receiver operating characteristics in recognition memory: evidence for a dual-process model', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.20, pp.1341–54.

Yonelinas, A.P. (1997) 'Recognition memory ROCs for item and associative memory: evidence for a dual-process signal detection model', *Memory and Cognition*, vol.25, pp.747–63.

Yonelinas, A.P. (2001) 'Consciousness, control and confidence: the 3 Cs of recognition memory', *Journal of Experimental Psychology: General*, vol.130, pp.361–79.

Yonelinas, A.P. (2002) 'Components of episodic memory: the contribution of recollection and familiarity', in Baddeley, A., Conway, M., and Aggleton, J. (eds.) *Episodic Memory: New Directions in Research*, Oxford, Oxford University Press.

Yonelinas, A.P. and Jacoby, L.L. (1996) 'Noncriterial recollection: familiarity as automatic irrelevant recollection', *Consciousness and Cognition*, vol.5, pp.131–41.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

CHAPTER 9

WORKING MEMORY

Graham J. Hitch

1 INTRODUCTION

Working memory (WM) refers to a central feature of cognition, namely our ability to process and maintain temporary information in a wide range of everyday activities such as planning, learning, reasoning, problem solving, and following instructions. This chapter begins by introducing the concept of working memory in the context of the distinction between short-term memory (STM) and long-term memory (LTM). Having done this, we are in a position to consider the structure of working memory, that is, the unchanging features that account for its operation in different activities. We shall see that – in common with many other aspects of the cognitive system – identifying structure is no trivial task. The discussion is organized around the multi-component model of working memory (Baddeley and Hitch, 1974; Baddeley, 1986), tracing some of the developments in this model in the light of new evidence and noting alternative accounts where appropriate. The material covered has been chosen to illustrate the increasing diversity of phenomena that are seen as relating to working memory and includes evidence from laboratory experiments, individual differences, normal and abnormal development, neuropsychology, and neuroimaging. We go on to focus in more detail on the particular topic of phonological working memory and vocabulary acquisition, where the convergence of different kinds of evidence is particularly striking. Finally, we take a short look at recent developments in computational modelling that make theories of some aspects of working memory more explicit and more precise. Overall, we shall see that the multi-component view of working memory provides a useful framework for explaining known facts and guiding further research, and can be used productively to address a range of practical problems.

1.1 Human memory as a multifaceted system

When someone tells us they have a poor memory, they may be referring to any of a range of specific problems. For example, they may have difficulties in recalling past events, remembering to do things, or perhaps retrieving facts or names. In everyday life we tend to talk about memory as if it were a single faculty. However, there are many grounds for thinking that memory is multifaceted, made up of a number of separate but interlinked systems (see Chapter 8). A long-standing theoretical distinction is that between LTM, a system for holding information over long periods of time, and STM (or primary memory), a system that holds information over much shorter intervals, as in looking up a reference number in order to reproduce it a few seconds later. Working memory is closely related to STM but goes beyond the mere retention of temporary information. More specifically, working memory not only maintains temporary information but supports operations on temporary information in a variety of cognitive tasks. A classic illustration is mental arithmetic, where we typically break a complex calculation into a series of simpler operations. For example, 26 + 37 might be broken down into the stages 20 + 30 = 50, then 6 + 7 = 13, and finally 50 + 13 = 63. The various stages have to be coordinated, and early stages generate transient information that has to be maintained for eventual use in later stages. Experimental studies show that errors of mental arithmetic are mainly due to forgetting temporary information during delays imposed by the sequencing of operations (Hitch, 1978), and one reason we often resort to written calculation is that having a durable external record bypasses some of the limitations of working memory. Other familiar examples of situations that place demands on working memory are interacting with a group of new acquaintances at the same time as trying to learn and remember their names, or taking notes while trying to follow a lecture. In such cases the combined demands of attending to mental operations while remembering transient information place a load on working memory and may result in errors. In order to discuss the constraints imposed by working memory in greater detail, it is necessary to sharpen the distinction between working memory and STM. This will be done in Section 1.3, but in order to get closer to the roots of this distinction, we need to first go back to the origins of the distinction between STM and LTM.

1.2 Distinction between short-term and long-term memory

William James introduced the concept of 'primary memory' in the late nineteenth century, but it was not until the 1960s that an interest in memory over brief intervals of less than a minute became firmly established. Memory researchers at that time became preoccupied with questioning the assumption that human memory is a unitary mental faculty. A number of different kinds of evidence were emerging to suggest that separate sub-systems of memory underpin immediate and delayed recall. One of them was evidence that memory for verbal stimuli has different properties over short and long intervals. For example, Baddeley (1966a) showed that immediate recall of a list of briefly presented words is poor when the items are phonemically similar (e.g. share the same vowel, as in *man*, *can*, *cad*, etc.) but is unaffected when they are semantically similar (e.g. share the same meaning, as in *huge*, *big*, *large*, etc.). However, when the same materials are presented more than once and memory is tested after a longer retention interval, the accuracy of recall is lower for semantically similar items and is unaffected by phonemic similarity (Baddeley, 1966b). These observations pointed to separate storage systems that code information in different ways, information in STM being held in an acoustic or speech-based code and information in LTM being coded in terms of meaning. Other evidence showed that the rate of forgetting briefly presented stimuli was unusually rapid when compared with forgetting rates for better-learned material, consistent with the idea that STM is much more labile than LTM (Brown, 1958). Over and above these observations, it had been known for some time that the 'span' of immediate memory, i.e. the longest sequence that can be recalled accurately after a single presentation, is limited to just a few digits, letters, or words (e.g. Miller, 1956). The limit on span suggested that STM can be distinguished from LTM on the grounds of its remarkably limited capacity.

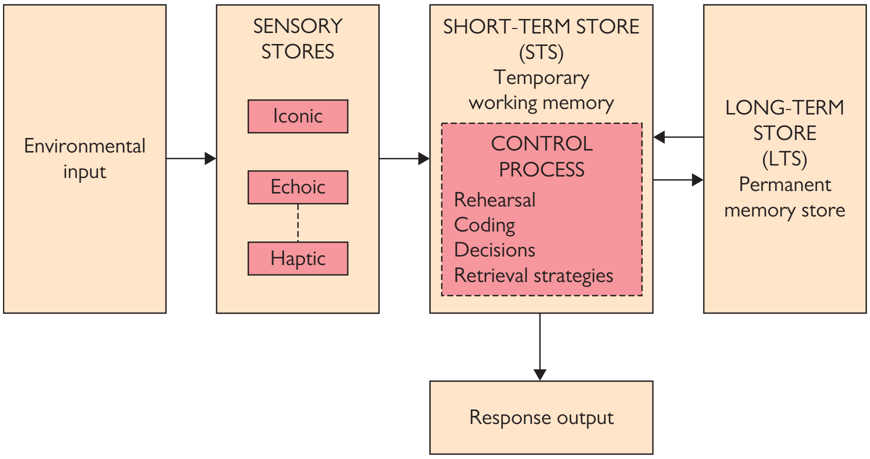


FIGURE 9.1 The modal model of memory, redrawn from Atkinson and Shiffrin (1971). Note how information has to pass through the short-term store in order to access the long-term store. Note also that the sensory stores are not discussed in the text. They are very short-lived and are specific to the various sensory pathways that feed information into the short-term store. *Source: Atkinson and Shiffrin, 1971*

So compelling was all this evidence at the time that several two-store models of memory were proposed. Reflecting this unanimity, their common features were referred to as the 'modal' model (Murdock, 1967). The main assumptions of this model were (1) that STM is a limited-capacity store subject to rapid forgetting, (2) that control processes, such as subvocal rehearsal, can be used to maintain information in STM, and (3) that information in STM is gradually transferred to LTM. Atkinson and Shiffrin (1971) provide the best known example of this type of account (see Figure 9.1).

You will see from the diagram that Atkinson and Shiffrin (1971) labelled their short-term store as a working memory that serves other functions besides acting as a temporary store. These functions include the regulation of control processes such as rehearsal and retrieving information from LTM. Note that control processes are volitional and are conceptually different from involuntary, automatic processes (such as becoming alert to the sound of one's own name whilst attending to something else). At about the same time that the Atkinson and Shiffrin model was enjoying its popularity, numerous other authors argued that the transient storage provided by STM was crucial for cognitive activities such as sentence comprehension and problem solving. In other words, there was a general assumption that STM behaves as some form of working memory. You can gain some insight into the plausibility of supposing that these activities require keeping track of temporary information within a stream of ongoing mental operations by trying one for yourself (see Box 9.1 on the comprehension of garden-path sentences).

Although the modal model captured some important insights, the consensus it reflected turned out to be fleeting. One concern was whether the various strands of evidence for distinguishing STM and LTM converged on a coherent account. For example, different ways of estimating the capacity of the short-term store gave quite different answers and reasons for this were unclear. A consequence of this particular challenge was a resurgence of interest in general principles of remembering as in the 'levels of processing' theoretical framework of Craik and Lockhart (1972). Another concern was whether it is appropriate to assume that STM behaves as a working memory. One difficulty for this position arose from intriguing neuropsychological evidence from a patient known in the literature as KF (Shallice and Warrington, 1970). KF sustained brain damage from an accident and his auditory digit span was reduced to two items, which is way below the normal range of about seven identified by Miller (1956). However, despite having such a low memory span,

BOX 9.1 Understanding 'garden-path' sentences

Garden-path sentences are sentences that lead the comprehender 'up the garden path' towards an incorrect interpretation, as in *We painted the wall with cracks* (see Chapter 6). It is the ambiguity of such sentences that makes them difficult. One explanation assumes that multiple interpretations of ambiguous sentences are held in working memory (Just and Carpenter, 1992). Just and Carpenter support their view with evidence that individuals with low working memory capacity are less able to maintain multiple interpretations than individuals with high working memory capacity. However, an alternative theory is that comprehension draws on more specialized resources than working memory (e.g. Caplan and Waters, 1999).

KF performed normally on tests of long-term learning and memory; he possessed normal intelligence and had no major difficulties in understanding spoken language (Shallice and Warrington, 1970). In general terms KF's pattern of memory performance was consistent with the modal model as it could be explained in terms of selective damage to STM while LTM was intact. Moreover, the fact that damage to part of the brain could have this effect suggested a separate neuroanatomical localization of STM. However, the absence of a general impairment in KF's learning, comprehension, and reasoning presents obvious difficulties for the idea that STM acts as a general-purpose working memory necessary for supporting a range of cognitive activities.

1.3 Working memory as more than STM

Given theoretical difficulties such as that presented by KF, Baddeley and Hitch (1974) made an empirical investigation of whether STM does indeed act as a working memory. One technique they used was the **dual-task paradigm** in which participants performed a STM task and a separate cognitive task at the same time. The logic of this paradigm is that two tasks will mutually interfere if they require access to a common resource and their combined demands exceed its capacity. Moreover, the effect of loading STM on cognition provides an experimental simulation of the depleted STM of neurological patients like KF. In a series of experiments, Baddeley and Hitch (1974) studied the effect of performing an irrelevant STM task at the same time as reasoning, comprehending language, or learning new information. For example, in one experiment people carried out a verbal reasoning task while remembering sequences of random digits (see Box 9.2 for an outline of the experimental procedure). Reasoning was impaired when the STM load was increased by making the digit sequences longer. Similar results were obtained when the cognitive task was either comprehending prose or learning a list of words for free recall. Baddeley and Hitch (1974) drew two main conclusions from these observations. First, the finding that an irrelevant STM task interferes with a range of cognitive tasks is consistent with the idea of a common working memory system that combines temporary information storage with ongoing mental operations. Second, while working memory overlaps with STM, the two concepts are not the same. Thus, even when the irrelevant digit load approached memory span and therefore 'filled' STM storage capacity, there was no catastrophic breakdown in any of the cognitive tasks. This suggests that working memory includes an additional resource not shared with STM. Baddeley and Hitch (1974) tentatively identified this as a 'central executive' responsible for the attentional control and coordination of mental operations.

Further evidence for a distinction between STM and working memory came from studies of individual differences. The logic behind this approach is that if two tasks involve similar underlying psychological processes, then a person who performs well on one should do well on the other. In statistical terms, the two abilities should be positively correlated. In an influential study, Daneman and Carpenter (1980) argued that

BOX 9.2 RESEARCH STUDY Studying the effect of an irrelevant memory load on verbal reasoning

The verbal reasoning task used by Baddeley and Hitch (1974) involved deciding whether a sentence gave a true or a false description of the order of a letter pair. Examples are: *A precedes B – AB* (true), and *B does not follow A – AB* (false). Varying the verb, the grammar, the letter order, and the truth-value of the answer gave a total of 32 problems of varying difficulty. Each problem was shown individually, performance being measured by the speed and accuracy of pressing 'true' and 'false' response keys.

One experiment involved a comparison between the effect on reasoning of concurrently repeating a sequence of six random digits and counting repeatedly from one to six. The rationale was that a sequence of six random digits is close to the span of immediate memory, whereas the counting sequence is stored in long-term memory. Repeating random digits slowed solution times in the reasoning task relative to a control condition, but the counting task had very little effect. Furthermore, the interference produced by random digits was greater for the more difficult versions of the reasoning task.

The conclusion Baddeley and Hitch (1974) drew was that reasoning and short-term retention compete for a limited-capacity 'workspace' that can be flexibly allocated either to the storage demands of the memory load or the processing demands of the reasoning task.

standard measures of STM, such as word span and digit span, tax storage capacity but do not assess the capacity to combine storage with ongoing processing operations. In order to provide a better assessment of the latter, and therefore of working memory capacity, Daneman and Carpenter devised a novel reading span task. In this task, participants were required to read aloud a set of unrelated sentences and immediately afterwards to recall the last word of each sentence. Box 9.3 gives further information about the procedure. To assess reading span, Daneman and Carpenter (1980) prepared sets of two, three, four, five, and six sentences. Participants were presented with progressively larger sets until they failed at a particular level. An individual's reading span was measured as the biggest set size at which he or she could achieve correct recall. Thus, reading span reflects how many items a person can remember at the same time as reading sentences, in contrast to their STM span, which reflects how many items they can remember when there is no requirement for simultaneous processing.

Daneman and Carpenter (1980) compared reading span with STM span for words as predictors of reading comprehension skills in a group of college students. Reading comprehension was measured in three ways: fact questions, pronoun questions, and verbal Scholastic Aptitude Test (SAT) scores (see Table 9.1). It turned out that reading span was a good predictor of all three measures of reading comprehension and a much better predictor than word span. Daneman and Carpenter went on to show that a listening span measure gave similar results, showing that the correlation is not specific to reading. They interpreted their findings as showing that working memory capacity is an important source of individual differences in language comprehension, the key characteristic of working memory being its function of combining temporary storage and information processing, in line with the approach taken by Baddeley and Hitch (1974).

A criticism made of Daneman and Carpenter was that their correlations might be an artefact of similarities in processing operations in the various tasks they used. Thus, reading span, listening span, and language comprehension all involve language processing, whereas word span does not. The potential force of this criticism is substantial and called into question whether Daneman and Carpenter's results have anything to do with working memory as a general-purpose resource. To address it, other investigators looked at patterns of correlation using different measures of working memory span to which the criticism does not apply. For example, Turner and Engle (1989) devised

BOX 9.3 RESEARCH STUDY Procedure for determining reading span

The materials for Daneman and Carpenter's (1980) reading span task were a set of unrelated sentences, each of which was typed on a separate card. The two examples they gave are:

Card 1: *When at last his eyes opened there was no gleam of triumph, no shade of anger.*

Card 2: *The taxi turned up Michigan Avenue where they had a clear view of the lake.*

Cards were arranged in sets of two, three, four, five, and six sentences, there being three instances of each set-size. Participants were shown one card at a time and read it aloud at their own pace, starting at set-size two. The second card was presented as soon as the first was read. A blank card signalled recall of the final word on each card in their order of occurrence (i.e. *anger, lake* in the above example of set-size two). Three trials were given at each set-size, and set-size was increased until all three trials at a particular level were failed. At this point testing was ended. Reading span was taken as the level at which the participant was correct on two out of three sets. As with memory span, there are many variants on this basic procedure.

TABLE 9.1 Correlations between spans and various measures of reading comprehension

	Reading comprehension measure		
	Fact questions	Pronoun questions	Verbal SAT
Reading span	.72	.90	.59
Word span	.37	.33	.35

Source: Daneman and Carpenter, 1980, Experiment 1

an operation span task in which participants solved a series of arithmetical calculations. A single word was presented after each calculation, and after completing each set of calculations, all the words had to be recalled. Operation span was measured as the maximum number of calculations for which this final recall was accurate. Turner and Engle (1989) found that operation span was a superior predictor of reading comprehension than standard STM span, despite involving dissimilar processing operations. Their results therefore support the idea of a general working memory system underpinning the ability to combine information processing with temporary storage in a range of different activities. Indeed, reading span, listening span, operation span, and tasks like them are often referred to by the generic term working memory span. Subsequent work has expanded the picture by showing that individual differences in working memory span predict performance in a range of tasks such as following directions (Engle *et al.*, 1991), are closely related to fluid intelligence (Engle *et al.*, 1999b), and can be used as longitudinal predictors of children's scholastic development (Hitch *et al.*, 2001).

SUMMARY OF SECTION 1

- Human memory is not a unitary system and appears instead to be multifaceted.
- An important distinction has been made between STM and LTM.
- Atkinson and Shiffrin (1971) suggested that STM acts as a working memory responsible for a variety of control processes.
- Baddeley and Hitch (1974) explored and expanded this idea and concluded that STM is better regarded as a component of working memory.
- Converging evidence that working memory differs from STM comes from studies of individual differences in tasks such as reading span and operation span.

2 THE STRUCTURE OF WORKING MEMORY

We have seen some of the evidence suggesting that working memory differs from STM, but so far little about how it differs beyond the suggestion that working memory involves an additional component, the central executive.

2.1 A multi-component model

In their original investigation, Baddeley and Hitch (1974) studied whether irrelevant STM loads affected reasoning, language comprehension, and list learning. Their aim was to examine whether these cognitive activities involve the same limited capacity as STM. Although high STM loads did cause interference, people could retain loads of two or three items without much disruption to the primary task. This observation was seen as consistent with the suggestion that working memory can be partitioned into two components, a store that can hold small amounts of temporary information and a central executive that coordinates mental operations. In further experiments Baddeley and Hitch (1974) looked at the effects of varying the phonemic similarity of the materials in reasoning and comprehension tasks. Adverse sensitivity to phonemic similarity is a characteristic feature of STM (see Section 1.2), and if reasoning and comprehension are sensitive too this would suggest that they share a common factor. In the reasoning task, subjects were asked to verify relationships such as 'A is not preceded by B – *AB*', where the letters used were either phonemically similar (e.g. TD) or dissimilar (e.g. MC). In the comprehension task, subjects were asked to say whether the words of a sentence were presented in a meaningful or jumbled order. The words either rhymed (e.g. *Red-headed Ned said Ted fed in bed*) or did not rhyme (e.g. *Dark-skinned Ian thought Harry ate in bed*). The results showed that phonemic similarity did indeed disrupt reasoning and comprehension, though only somewhat mildly.

To account for their results, Baddeley and Hitch (1974) assumed that one of the components of working memory is a limited-capacity, speech-based store capable of storing two to three items. This sub-system was described as the **articulatory loop** (subsequently renamed the phonological loop) and can be viewed as roughly equivalent to the earlier concept of STM (more detail about the phonological loop is given in Section 2.2). The phonological loop could be used to store small memory loads during cognitive tasks and was responsible for the effect of phonemic similarity on performance. The second component was the **central executive**, responsible for the control and coordination of mental operations in a range of activities including but extending beyond reasoning, comprehension, learning, and memory. The executive was seen as a limited-capacity workspace that can be flexibly allocated to control processes or temporary information storage, depending on the nature of the task in hand. Thus a small irrelevant memory load could be stored in the phonological loop without taxing the central executive, but a larger memory load would take up extra resources in the executive. Given a limit on the capacity of the workspace, this theoretical account maintains that there will be a trade-off such that fewer resources are available to support processing operations when temporary storage demands increase.

Reflecting on their results, Baddeley and Hitch noted that the tasks they had investigated were all primarily verbal. The question arose as to whether tasks involving visual memory and visual imagery also draw on working memory and, if so, how. The information available from dual-task studies indicated that combining two visuo-spatial activities (such as tracking a moving object while performing a mental imagery task) or combining two verbal activities is more difficult than combining a visuo-spatial and a verbal task. This observation suggests that there are separate resources specialized for dealing with verbal and visuo-spatial information. Nevertheless, as there is some mutual interference when a visuo-spatial and a verbal task are combined, the data are also consistent with the involvement of a common resource. One way of accounting for these observations is to assume that the central executive controls visual and verbal tasks and that there is a separate sub-system for storing visuo-spatial information, analogous to the phonological loop. This tripartite model, in which the extra sub-system is referred to as the visuo-spatial sketchpad, was developed further by Baddeley (1983; 1986) and is illustrated in Figure 9.2.

Neuropsychological patients can show selective impairments in either visuo-spatial or verbal working memory. For example, de Renzi and Nichelli (1975) found Corsi span and auditory digit span could be independently impaired in patients with different types of lesion. Corsi span is a test of visuo-spatial STM that uses an apparatus in which a set of nine identical cubes is placed at haphazard locations on a horizontal board. The experimenter points to a selection of cubes one by one and the task is to reproduce the sequence immediately afterwards by pointing. As in digit span, sequence length is progressively increased to find the limit where performance breaks down. The observation of a double dissociation of neuropsychological impairment is strong evidence for separate verbal and visuo-spatial stores.

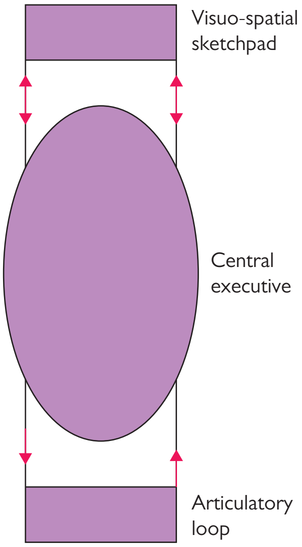


FIGURE 9.2 The structure of working memory. *Source: based on Baddeley, 1983*

Dual-task methods have proved to be a highly productive way of investigating the nature of the visuo-spatial sketchpad. In one highly cited study, Baddeley and Lieberman (1980) found that the ability to use a mental imagery mnemonic to remember verbal materials was disrupted by a concurrent spatial, non-visual task (tracking a moving loudspeaker while blindfold) but not by a concurrent visual, non-spatial task (detecting changes in the brightness of a blank field). This pattern was not observed when participants used rote rehearsal rather than imagery, suggesting it was not a function of the relative difficulties of the interfering tasks. Baddeley and Lieberman (1980) explained their results in terms of the fact that the imagery mnemonic they studied was spatial rather than visual. Subsequent research has largely borne out the idea that the visuo-spatial sketchpad has separate spatial and visual components (Klauer and Zhao, 2004; Logie, 1995).

An interesting example of a predominantly visual imagery task is provided by the work of Hitch *et al.* (1995). Participants were first shown two line drawings, one after the other, and were then asked to superimpose them using mental imagery and identify the novel figure thus revealed. For instance, in one trial the first drawing showed two ice cream cones placed apart from one another horizontally. The second drawing showed a line shaped as an inverted U with ends where the tops of the cones had appeared just previously. When mentally superimposed, the two drawings combined to reveal a skipping rope with the cones as the handles. Hitch *et al.* (1995) found that identifying the novel figure was more accurate when both drawings consisted of a black figure on a white ground than when one was white on black, providing clear evidence that the images preserved information about visual appearance. It is interesting to note in passing that if you were able to 'see' the skipping rope in your mind's eye in the above example, you achieved this using conceptually-driven images rather than the perceptually-driven images studied by Hitch *et al.* (1995). This is an important distinction as the characteristics of these two types of image are not necessarily the same (see for example Farah, 1984).

In reviewing visuo-spatial working memory, Logie (1995) suggested that the spatial and visual subsystems comprise a spatial movement system that can be used to refresh the contents of the visual store, by analogy with the role of articulatory rehearsal in the phonological loop. Smyth and Waller (1998) asked rock climbers to imagine tackling familiar routes while performing a variety of secondary tasks designed to disrupt their ability to use visual, spatial, or kinaesthetic information and obtained results implicating all three dimensions of representation. However, we are still some way from understanding how these different types of information combine together.

In conclusion, the work of Baddeley and Hitch (1974) led to a tripartite model of working memory that was subsequently developed by Baddeley (1986). This model appears to have been the first substantive account of working memory and has been influential within the field. However, an increasing number of alternative accounts has emerged subsequently, many of which are described in a landmark volume edited by Miyake and Shah (1999). Several theoretical issues divide these approaches. One of the main ones concerns the relationship between working memory and LTM. Baddeley and Hitch (1974) assumed separate systems. However, a number of authors take a different view, maintaining that working memory corresponds to the currently activated sub-region of LTM (e.g. Ericsson and Kintsch, 1995; Cowan, 1988; 2005). Part of the motivation for this alternative approach comes from evidence that extensive knowledge of a specific domain is associated with a dramatic increase in working memory capacity for information from that domain. For example, chess experts display superior working memory skills when reproducing an arrangement of chess pieces on a board when the arrangement is plausible in the context of a game. However, with a random arrangement of pieces experts perform no better than non-experts, suggesting that the underlying capacity of working memory is the same in both groups and that experts benefit from their greater familiarity with patterns taken from games. If we think of the capacity of working memory in terms of chunks, where each chunk is a unit corresponding to a familiar subgroup of items, then we can understand these results in terms of experts being able to encode familiar stimuli as larger chunks when presented with plausible arrangements. Interestingly, Cowan (1988; 2005) maintains the assumption that working memory involves a limited capacity attentional system (see also Engle *et al.*, 1999a). The idea of specialized buffer stores has also been challenged by the work of Jones (see Section 2.3.4). In the remainder of this chapter we stay within the Baddeley and Hitch (1974) framework for the purpose of organizing the discussion, raising problems for it where appropriate. We begin with the relatively well-specified concept of the phonological loop, before moving on to the central executive, the most important but least understood aspect of working memory.

2.2 Phonological working memory

One reason the phonological loop is relatively well specified is the existence of a cluster of interrelated experimental manipulations that affect its operation. We have already encountered one of these, namely the phonemic similarity of items presented in tests of immediate recall (see Section 1.2). A second variable was the word length of the items. In an important series of experiments, Baddeley *et al.* (1975) showed that the limit on STM span for verbal stimuli was not simply a fixed number of items or chunks, as Miller (1956) had claimed, and in particular was higher for shorter words (e.g. *harm*, *wit*) than for longer words (e.g. *university*, *hippopotamus*). Box 9.4 describes one of their procedures and its results. One of many interesting observations was a systematic relationship between how many words could be recalled and the time it took to say them out loud. Thus, people could recall the number of words that could be spoken in about two seconds. This is consistent with the idea of a rehearsal loop in which rehearsing items sequentially refreshes their decaying memory traces. Longer

BOX 9.4 RESEARCH STUDY The word-length effect

In one of their experiments, Baddeley *et al.* (1975) constructed five pools of 10 words of one, two, three, four, or five syllables. The pools were matched for semantic category and familiarity. To illustrate, the one-syllable pool included *stoat, mumps, school, Greece*, and corresponding items in the five-syllable pool were *hippopotamus, tuberculosis, university, Yugoslavia*. Ten lists of five words were made up of random permutations within each pool. The lists were presented in a random order, words being shown one after another at a two-second rate. Immediately after list presentation, participants spoke their recall. In a second part of the experiment, reading rate was measured. This was achieved by timing participants reading aloud a typed list of the words in each pool as quickly as they could.

The results showed that the percentage of words recalled dropped as the number of syllables increased. Moreover, as the graph shows (see Figure 9.3), the plot of percentage correct recall against articulation rate formed a straight line. The slope of the line indicates that the phonological loop has a time-based capacity such that people can recall as many words as they can say aloud in about two seconds. Individual differences paint a corresponding picture in that the faster a person can say a list of words out loud (that is, the faster they can rehearse), the more effective they prove in subsequently recalling those words.

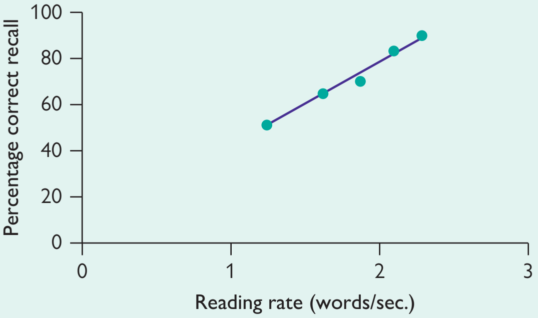


FIGURE 9.3 Results obtained by Baddeley *et al.* (1975). Percentage of words recalled is plotted as a function of the rate at which the same words could be read aloud, for five different word-lengths. The point furthest to the right corresponds to one-syllable words, the next point to the left represents two-syllable words, and so on.

words take longer to rehearse and so fewer can be refreshed within the time limit set by the rapidity of the decay process. Baddeley *et al.* (1975) also examined individual differences and found that faster speakers tended to recall more information than slower speakers. This is consistent with the model given the plausible assumption that faster speakers can rehearse more rapidly. The model could also account for the phonemic similarity effect, as a given amount of decay would have a greater effect on the ability to discriminate the memory traces of items that share phonological features. To appreciate this point, suppose you have been presented with the sequence of phonemically similar letters BTCG to recall. If as a result of partial forgetting you could only remember that the third item contained an /e/ sound, this would not be very helpful as it leaves many options open. Compare this with a sequence of dissimilar items such as *RJQL*, where being able to remember that the third item had a /u/ sound would be of much more help.

This model of the phonological loop was also able to explain the results of dual-task experiments in which immediate serial recall was combined with articulatory suppression (a secondary task involving the repetition of a redundant and irrelevant word such as *the the the the*). This low-level secondary task is intended to occupy the phonological loop with irrelevant (but unavoidable) activity, so that performance on the primary task has to manage without the assistance of the phonological loop (or at least without a large part of its functioning). Baddeley *et al.* (1975) found that articulatory suppression disrupted recall, consistent with disrupting the phonological loop.

Suppression also removed differences between the recall of longer and shorter words and between phonemically similar and dissimilar items. These further effects are also consistent with disruption of the loop. However, the effects of word length and phonemic similarity only disappeared when items were presented visually and not when they were presented auditorily. This unexpected effect of presentation modality was for some time something of a puzzle. The position was eventually clarified in experiments carried out by Baddeley *et al.* (1984) where suppression was continued during recall as well as item presentation. Under these conditions, suppression removed the word-length effect for auditory items, but still did not remove the phonemic similarity effect. Baddeley *et al.* (1984) explained these results in terms of a modified theoretical account in which the phonological loop consists of a decaying phonological store (the locus of the phonemic similarity effect) and a control process of sub-vocal rehearsal (the locus of the word-length effect) (see Figure 9.4). According to this account, spoken stimuli access the loop automatically whereas visual inputs have to be verbally recoded, an optional control process that involves **sub-vocalization**. Suppression eliminates the word-length effect for both visually and auditorily-presented stimuli by disrupting rehearsal, but only eliminates the effect of phonemic similarity for visually-presented stimuli, as only this method of presentation requires verbal recoding. In this way, specification of different pathways by which visual and spoken stimuli access the loop explains an otherwise obscure pattern of findings. The next section shows how this model of the phonological loop generates useful insights into developmental changes in verbal STM as children grow up.

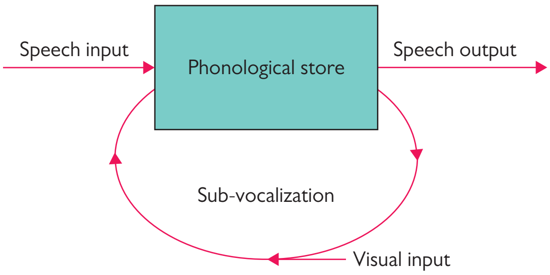


FIGURE 9.4 The structure of the phonological loop, according to the ideas developed by Baddeley *et al.*, 1984.

2.2.1 Developmental and cross-linguistic differences

The two-part model of the phonological loop is interesting in a number of different ways. Not least is that the model can be applied to phenomena outside its initial scope. One example is the developmental growth of memory span during childhood, for which many competing explanations had been proposed (Dempster, 1981). Thinking in terms of the phonological loop suggests it would be informative to measure children's recall of lists of words of different lengths and the speed at which they can articulate the words, as in Baddeley *et al.*'s (1975) study of adults. The results of doing this are quite striking. As children's ages increase, their average level of recall increases in proportion to the rise in their average speech rate (Nicolson, 1981; Hulme *et al.*, 1984; see Figure 9.5). Furthermore, the size of the word-length effect in children of different ages reflects the time it takes to articulate words of different lengths. Finding

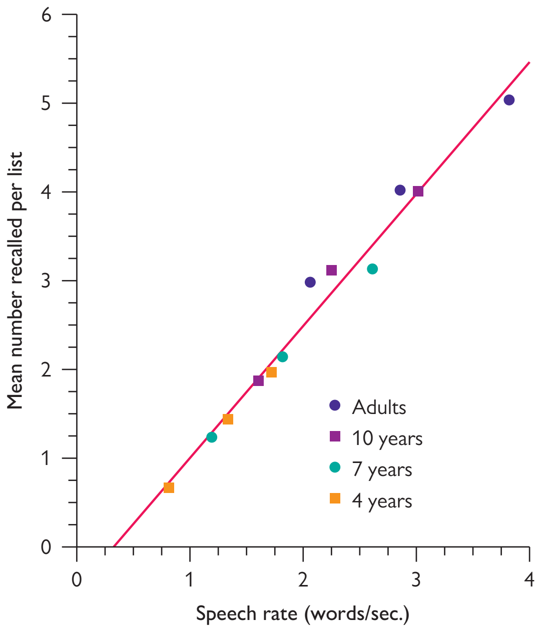


FIGURE 9.5 Results obtained by Hulme *et al.*, 1984. Percentage of words recalled is plotted as a function of their speech rate for three different word-lengths and four age groups.

such a clear empirical relationship is informative and suggests a possible explanation for the developmental growth in memory span, namely that older children can rehearse faster and can therefore maintain more items within the time-limit set by trace decay in the phonological store. The phonological loop suggests a similar explanation for the existence of some curious cross-linguistic differences in STM span. For example, differences in mean digit span between English (7.2), Spanish (6.4), Hebrew (6.5), and Arabic (5.8) are highly correlated with the rates at which the digits can be articulated in these languages, faster rates being associated with higher spans (Naveh-Benjamin and Ayres, 1986).

The phonological loop model has prompted further questions about developmental change. One of these concerns how children remember a series of nameable objects. Older children and adults show the standard tendency for poorer recall of items with longer names. However, children under the age of seven show no such effect, though they do show a word-length effect when the items are presented in the form of their spoken names instead of pictures (Hitch *et al.*, 1989). Moreover, when recalling nameable pictures, younger children find it harder when the pictures are visually similar to one another, whereas older children find it harder when the names of the pictures are phonemically similar (Hitch *et al.*, 1988). These observations are consistent with the assumption that spoken stimuli gain automatic access to the loop but that phonological recoding is necessary for visual stimuli (see Section 2.2). They suggest further that the process of recoding is somewhat slow to develop and that younger children are more reliant on visuo-spatial working memory. Subsequent research has confirmed the developmental progression from visual to phonological coding and suggests that it may be related to learning to read, being markedly delayed in dyslexic children (Palmer, 2000a; 2000b).

2.2.2 The irrelevant speech effect

Yet another application of the phonological loop model was to explain why exposure to background speech sounds disrupts STM for visually presented verbal stimuli. Salamé and Baddeley (1982) showed that having to ignore irrelevant speech was more interfering than ignoring irrelevant noise, leading them to suggest that unattended speech enters the phonological store whereas non-speech sounds do not. Consistent with such an interpretation, blocking people's ability to verbally recode visual stimuli by having them perform articulatory suppression removes the disruptive effect of irrelevant speech (Salamé and Baddeley, 1982). However, the phonological loop interpretation was challenged by evidence that irrelevant non-speech sounds can interfere with verbal STM, the amount of interference being determined by the same factors as for speech. Thus steady-state streams, where the irrelevant sounds remain the same, cause less disruption than changing-state streams, where the irrelevant sounds vary over time (Macken and Jones, 1995). Such observations have been used to question the assumption that irrelevant speech has an effect that is specific to the phonological loop. They suggest a broader explanation of the interference caused by irrelevant sounds in terms of general memory mechanisms that are not specific to the verbal domain, as in Jones's (1993) concept of a multimodal 'episodic record'. The explanation of irrelevant sound effects has developed into an area of considerable controversy (see for example Baddeley, 2007, Chapter 3; Baddeley and Larsen, 2003; Jones *et al.*, 2006). Although this debate continues, a possible resolution is that irrelevant speech affects a serial ordering mechanism common to speech and non-speech sounds (see for example Beaman and Jones, 1997) as well as the speech-specific phonological store (Page and Norris, 2003).

2.2.3 Neural basis

Before closing this part of the discussion, we shall consider some evidence about the neural basis of the phonological loop. An obvious challenge for any model is to explain neuropsychological cases of selective impairment of memory span of the type demonstrated by Shallice and Warrington (1970). Vallar and Baddeley (1984) made a detailed investigation of one such patient, known as PV. Following a stroke damaging the left parietal cortex, PV's auditory digit span was reduced to only two items. However, many of her other abilities were unimpaired. For example, her speech was fluent and her rate of articulation was normal. Vallar and Baddeley (1984) found that PV's memory span for spoken sequences was poorer for phonemically similar items but unaffected by items' word length, which they interpreted as indicating a phonological store deficit. If the store was functioning at a reduced level, spoken inputs would nevertheless access it automatically and immediate recall would be sensitive to phonemic similarity. However, PV would not find sub-vocal rehearsal a useful strategy with a damaged store and hence would not show an effect of word length. Like other patients of this type, PV's memory span for verbal materials was higher with visual than spoken presentation. Moreover, her visual span was unaffected by phonemic similarity or word length of the materials. PV may have been relying on visuo-spatial working memory to remember visual stimuli and there may be an interesting parallel to be drawn with children's reliance on visuo-spatial working memory for remembering visual stimuli early on in development when, albeit for different reasons, their ability to use the phonological loop is restricted (see Section 2.2.1).

The advent of neuroimaging techniques provided the opportunity to study the neuroanatomical basis of the phonological loop in the normal brain. In an early study, Paulesu *et al.* (1993) investigated which areas of the brain are active in tasks thought to involve the phonological loop. Such experiments depend on a subtraction logic whereby brain activation observed in one experimental task is compared with that in another. By arranging that the two tasks differ solely in the process of interest, the areas of activation specific to that particular process can be obtained by subtraction. This of course is not as simple as it sounds and typically involves making theoretical assumptions about the tasks under consideration. Paulesu *et al.* (1993) compared activation patterns in a verbal memory task requiring storage and rehearsal, a rhyme judgement task that required rehearsal but not storage, and a control task requiring neither storage nor rehearsal (Box 9.5 describes the experiment in more detail). The results suggested separate localization of storage and rehearsal, consistent with the theoretical distinction between these two aspects of the phonological loop. Furthermore, localization of the store to an area in the left parietal cortex corresponded approximately to the locus of damage in patients like PV. Other neuroimaging studies of working memory converge with – but also complicate – this simple picture, especially with regard to the involvement of other brain areas (see for example Henson, 2001; Baddeley, 2007, Chapter 13). Our purpose here is merely to illustrate an early success in using the phonological loop model to guide the collection and interpretation of neuroimaging data.

2.2.4 Theoretical issues

We have seen how a relatively simple model of the phonological loop has been productive in ways that extend well beyond its initial remit. These applications cover a surprisingly extensive range that includes developmental and cross-linguistic differences, effects of irrelevant speech, cases of neuropsychological impairment, and results of neuroimaging studies. The model has turned out to be remarkably successful – it has evidently 'travelled well'. However, some of its limitations are steadily becoming more apparent, as in its explanation of the effects of irrelevant speech. Other recent evidence suggests that the word-length effect may not be due to differences in items' spoken duration. Thus, there is little or no effect of word duration when the phonological complexity of items is carefully controlled (Lovatt *et al.*, 2000). Another difficulty is that young children show a clear word-length effect in STM for spoken items before developing the ability to use sub-vocal rehearsal (Gathercole *et al.*, 1994; Hulme *et al.*, 1984). Other authors have shown that output delays are sufficient to cause word-length effects, without appealing to rehearsal (Brown and Hulme, 1995; Cowan *et al.*, 1992; Gathercole and Hitch, 1993).

Whether the limitations of the phonological loop as a model count as falsifications is an interesting scientific issue that could send us back to the drawing board for an entirely new account. Some authors have taken this approach (Nairne, 2002). The alternative strategy is to revise the model to overcome its limitations, while at the same time preserving its original insights. We saw an earlier example of this in the elaboration of

BOX 9.5 RESEARCH STUDY Neural correlates of the phonological loop

Paulesu *et al.* (1993) used positron emission tomography (PET) to measure blood flow in different regions of the brain. This technique involves giving an intravenous injection of radioactive water and then scanning the brain to record the spatial distribution of radioactivity. Scanning is performed during matched tasks that differ with regard to a feature of interest. Subsequent comparison of the two activation patterns allows brain regions associated with the feature of interest to be identified. (A similar logic applies to functional magnetic resonance imaging (fMRI), a more recent technique that does not involve radioactivity.)

Paulesu *et al.* compared brain activation patterns in phonological and non-phonological memory tasks. The phonological task involved showing a sequence of six consonants followed by a probe letter. Participants indicated whether the probe item had appeared in the sequence. The non-phonological memory task was identical except that the items were unfamiliar Korean characters. The two tasks were therefore closely matched, but only remembering consonants engaged the phonological loop. Subtracting activation patterns revealed that the consonant memory task was associated with increased blood flow in left hemisphere regions corresponding to Broca's area and the supramarginal gyrus of the parietal cortex.

A second comparison was between a rhyme judgement task and a shape judgement task. In the rhyme task participants saw a series of consonants and indicated whether each one rhymed with the letter B, which was always present. The shape task was identical except that the stimuli were Korean characters and the judgement was one of shape similarity. Previous research suggested that the rhyme judgement task would engage the sub-vocal rehearsal system but not the phonological store. Subtraction of the scans indicated that the rhyme task activated Broca's area but not the left supramarginal gyrus. Thus, the sub-vocal rehearsal system can be identified with Broca's area and, by revisiting the subtraction for the memory tasks, the phonological store can be identified with the left supramarginal gyrus.

the account of the phonological loop to explain why the effects of articulatory suppression differ when the memory items are seen rather than heard (see Section 2.2). A more recent example is the effort to develop the concept of the phonological loop through more detailed computational modelling (see Section 4). It is probably too soon to say which of these strategies will be the more productive – a totally new approach or development based on the present model. Only time will tell. For the present we note that, despite its limitations, the phonological loop continues to provide a simple, usable framework for linking a robust set of psychological phenomena (see for example Baddeley, 2007, Chapter 3) and is still widely used. However, before continuing with further discussion of the phonological loop, we turn to the main aspect of working memory in the tripartite model: the central executive.

2.3 Executive processes

The central executive is, in general terms, responsible for controlling and coordinating mental operations in working memory. Baddeley and Hitch (1974) suggested that the functions of the executive included supervising slave stores such as the phonological loop and the visuo-spatial sketchpad, as well as interactions with LTM. However, as we shall see, more precise identification of executive functions is a matter of continuing development and debate. The executive is at once the most important component of working memory, the most controversial, and the least understood. At various times it has been described as a 'rag-bag' or an area of 'residual ignorance' and, in a review, Andrade (2001) referred to it as 'problematic'. There are good reasons for these remarks. One is that the executive could be seen as a reinvention of the homunculus, a person inside the head. The well-known problem here is that of explaining what controls the homunculus without appealing to an infinite regress of further homunculi. Another difficulty is that, at an intuitive level, executive processes clearly have links to our sense of conscious awareness. This is another difficult concept with a long history of intractability (see Chapter 18 on consciousness; see also Baddeley, 2007, Chapter 16). Fortunately, rather than allowing themselves to be put off by these problems, researchers have attempted to understand what aspects of executive control they can, with the long-term goal of steadily reducing the area of residual ignorance.

2.3.1 Central workspace

We read earlier how Baddeley and Hitch (1974) conceptualized the executive as a limited-capacity central workspace with resources that could be flexibly allocated to various combinations of mental operations and temporary information storage. We also saw how Daneman and Carpenter's (1980) reading span task was designed as a method for assessing the capacity of such a workspace. Thus, given the assumption that resources for processing and storage trade off against each other, reading span can be interpreted as a measure of residual storage capacity when the workspace is also occupied in supporting reading processes. A major boost for this hypothesis came from the work of Case *et al.* (1982) on counting span, a task with the same structure as reading span but a different combination of processing and storage. Participants count the number of target items in a series of visual displays, each containing a mixture of targets and distractors, and at the end of the set they attempt to recall the count totals. Like reading span, counting span is determined as the set size above which accurate recall is no longer possible. Case *et al.* (1982) found that the observed growth of counting span during childhood could be predicted from an observed increase in the speed of counting. Furthermore, when counting was artificially slowed down in adults by requiring them to use unfamiliar nonsense syllables instead of normal number names, this reduced their counting span to the level predicted by their slower counting speed. Case *et al.* (1982) took these results as evidence for a simple model in which speed is a marker for the proportion of limited workspace required to support counting, and hence for the remaining proportion free for storing count totals. However, further evidence is needed to confirm that it is useful to think of the span of working memory as reflecting the capacity of a central workspace or 'mental blackboard'.

Several investigators have tried to examine more precisely what limits the span of working memory in tasks such as reading span and listening span. Given that the number of items in store increases from the start to the end of a trial, the workspace hypothesis predicts a decline in the resources available to support processing as each trial progresses. This would follow from the trade-off between resources within the workspace. Towse *et al.* (1998) tested this prediction by studying the performance of children on reading span, operation span, and counting span in a series of parallel experiments. The results gave no clear support for a resource trade-off, in that there was no systematic change in the speed of processing operations within trials. Towse *et al.* (1998) entertained an alternative hypothesis according to which, rather than sharing attention between processing and storage, children switch attention back and forth between processing and storage. Thus, in reading span for example, children might read a sentence, store the final word, read the next sentence, store its final word, and so on. According to this 'task-switching' account, reading span is limited by the rate of forgetting sentence-final words during the time intervals spent in reading. This is similar to the way in which errors in mental arithmetic were explained (Hitch, 1978) and is quite different from the resource-sharing account. To test the task-switching hypothesis, Towse *et al.* (1998) manipulated the time intervals over which information had to be stored in different conditions in which the total amount of processing was held constant. This was achieved by altering the order of presentation of the items within a set, some of the items being designed to take longer to process than others. In line with the prediction from task-switching, spans were lower when the intervals over which information had to be maintained were longer. This was true for all three tasks, reading span, operation span, and counting span, suggesting a result of some generality. Subsequent research confirmed this by showing that manipulating the order of presentation of items has similar effects in adults (Towse *et al.*, 2000).

Other investigators have also found an effect of the length of the intervals devoted to processing operations in working memory span tasks, but have also shown that span is lower when the operations themselves are more complex (Barrouillet and Camos, 2001). Effects such as these have led towards a mixed model that involves both attention switching and resource sharing. Further evidence suggests that other factors are also involved in limiting working memory span. For example, de Beni *et al.* (1998) found that individuals with low working memory spans made more intrusion errors of recalling items from previous trials, suggesting that the ability to inhibit potentially interfering information is an important aspect of working memory capacity (see also Conway and Engle, 1994). To complicate matters further, it seems that the capacity to inhibit breaks down into two or three distinct subtypes of inhibition (Friedman and Miyake, 2004).

Taking all these observations together, it seems that a simplistic interpretation of working memory span as reflecting the capacity of a central workspace is unlikely to be entirely accurate. Working memory span tasks clearly share something important in common but seem nevertheless to require a more complex theoretical account. Such a conclusion points to the difficulty of sustaining any simple conceptualization of executive processes. Indeed, an important issue to emerge in recent studies of executive function is whether the executive is a single, unified entity (e.g. Duncan and Owen, 2000) or a system that is fractionated into distinct sub-components (e.g. Shallice, 2002). This question of fractionation has led to an interest in tasks other than working memory span that capture different aspects of executive function.

2.3.2 Attention

The view of the executive put forward by Baddeley (1986) was substantially different from that proposed earlier by Baddeley and Hitch (1974), stemming in part from difficulties with the idea of resource tradeoff. It was inspired by an imaginative attempt of Norman and Shallice (1986) to provide a unified explanation for slips of action in everyday life and the more serious disturbances of behaviour seen in patients with frontal lesions (frontal patients). One striking example of such a disturbance is 'utilization behaviour' (Lhermitte, 1983) where patients have difficulty inhibiting stereotyped responses to everyday objects. For instance, the sight of a toothbrush during a clinical assessment may lead to a patient brushing his or her teeth even though this is totally inappropriate to the situation.

Norman and Shallice (1986) proposed a model in which the control of action involves two levels. At the lower level is a set of learned schemata for routine sequences of actions or mental operations, each of which fires automatically to a specific 'trigger stimulus'. For example, if we overhear someone mention our own name we automatically orient our attention towards the speaker. These schemata are arranged in parallel, so that at any moment there is competition among those that potentially might fire. At the higher level sits a supervisory attentional system (SAS), a limited-capacity resource capable of intervening at the lower level. A typical example would be the SAS intervening to stop a schema from firing despite the presence of its trigger stimulus. The Norman and Shallice model explains the difficulties of frontal patients in terms of a deficit in the resources available for executive control. Thus in utilization behaviour, strongly triggered schemata fire even when they lead to contextually inappropriate behaviour. Diary studies of slips and lapses in everyday life reveal that these too often involve making an inappropriate but frequent action in a familiar context. For example, one diarist recorded intending to get his car out of the garage, but on the way there he stopped to put on his boots and gardening jacket, as if to work in the garden (Reason, 1984). Such 'capture errors' tended to occur at choice points when diarists reported their attention being distracted elsewhere. The Norman and Shallice (1986) model would explain such errors in terms of distraction rendering the SAS temporarily unavailable to inhibit a familiar and therefore strongly triggered habit at the choice point.

Baddeley (1986) adopted the SAS as a model of executive control, thus moving away from the notion

BOX 9.6 METHODS Random generation

In the random generation task, participants are asked to select items repeatedly at random from a restricted pool such as the digits 0–9 or the letters of the alphabet. Generation is usually required at a specified rate, such as one per second. Some idea of the difficulty of the task can be gained by asking someone to try it for a minute and noting down their responses. Most people soon start hesitating or repeating themselves, typically emitting stereotypical sequences such as alphabetic runs (e.g. *ABC*) or familiar acronyms (e.g. *FBI*). The degree of randomness can be estimated in various ways, one of the simplest being to count the proportion of stereotyped pairs produced. Baddeley (1986) described evidence that randomness declined systematically when either the pace of generation or the difficulty of a secondary card-sorting task was increased. These observations are consistent with the suggestion that random generation taxes a limited-capacity system.

of the executive as a workspace combining both processing and storage, to a purely attentional system. This theoretical move led more or less directly to a search for fresh ways of investigating executive processes. One such task involves generating a random stream of responses using only the digits 0–9, a surprisingly difficult task (see Box 9.6). A major source of difficulty is the avoidance of stereotyped sequences such as ascending or descending series of digits, or in the case of letters, alphabetical runs. This type of error is consistent with a task analysis in which the requirement for randomness involves pitting the capacity for supervised inhibitory control against the tendency to execute strongly learned habits, sometimes called 'pre-potent' responses. Experimental evidence confirms that random generation is a demanding task, but also shows that it is a very complex task, suggesting that it is unlikely to be a pure measure of executive function (see Towse, 1998).

2.3.3 Fractionation

In an attempt to develop the concept of the executive still further, Baddeley (1996) proposed that the system could be fractionated into a number of separate but related functions dealing with different aspects of attention. These were focusing, dividing, and switching attention, and using attention to access information in LTM. To give a general idea of these distinctions, focusing is required when irrelevant information has to be ignored, whereas dividing is necessary when attention has to be shared between different tasks. Thus attention is focused when listening to one message and ignoring another, but divided when two messages have to be monitored simultaneously, or when different activities have to be combined, as in dual-task experiments. Attention-switching, on the other hand, refers to situations where attention must be repeatedly shifted from one process to another. For example, in generating a random sequence of digits, attention must constantly shift between different retrieval plans in order to avoid stereotypical patterns of responses. This in turn is somewhat different from the role of attention when actively searching for information in LTM. Baddeley (1996) described a certain amount of empirical support for the separability of executive functions. For example, patients with Alzheimer's disease have an exaggerated difficulty in combining concurrent tasks, whereas normal ageing is associated with increasing difficulty in focusing attention. However, in general the paper was theoretical and was in essence an attempt to set the agenda for future research.

One way that the agenda has been taken forward is through the study of individual differences in executive function in the normal population. In one such study, Miyake *et al.* (2000) gave a large sample of students a range of tasks designed to involve different facets of attentional control. These were shifting attention, monitoring and updating information, and inhibiting pre-potent responses. Analysis of the data showed that a statistical model based on these three components gave a better account of relationships among abilities than simpler (i.e. one or two-factor) models. This outcome is broadly consistent with the general idea that executive function is fractionated, but the number of functions and their identity differ from Baddeley's (1996) proposal. Such a discrepancy is difficult to interpret, especially as a limitation of factor analysis is that it can only reveal the structure in the variables that are entered into the analysis. Miyake *et al.* (2000) went on to assess individual differences in a number of other tasks that are widely used as tests of executive function. The results showed that these tasks mapped onto the three putative components of executive function in different and sometimes unexpected ways. These are interesting findings and illustrate the need for further development in measuring executive function (see also Section 2.3.1).

As a general conclusion, the present state of knowledge is that executive function appears to fractionate, but it is not yet clear how (compare this with Chapter 2 on multiple types of attention). Thus, we still need to separate out and identify the various components of executive control (see Baddeley, 2007, Chapter 7). Whatever the outcome, there is a further issue of how such a diverse executive can operate in a unitary way. That is, how do the components of a many-faceted executive system interact coherently and avoid conflict in the control of perception, thought, and action?

2.3.4 Coherence and the binding problem

It is interesting to note that the problem of coherence is not restricted to executive processes and applies to working memory more generally. Thus, if any system consists of a number of separate sub-systems, then the question arises as to how the sub-systems interact to ensure that the system as a whole operates in an integrated manner. For example, if visuo-spatial information about multiple objects is stored separately from verbal information about the same objects, then the system must have a way of keeping track of which information refers to what object. This is sometimes referred to as the **binding problem**. Indeed, one critique of the working memory model of Baddeley and Hitch (1974) and its subsequent development by Baddeley (1986) is that the assumption of separate sub-systems creates a binding problem that the model fails to address (Jones, 1993). We encountered Jones's work when discussing the disruptive effect of irrelevant sounds on immediate memory for verbal sequences (Section 2.2.2 and Chapter 2). Jones *et al.* (1995) argued that interference from speech and non-speech sounds obeys the same principles, even when the memory task involves spatial information, and is best explained in terms of a common level of representation in a unitary memory system. This common 'episodic record' minimizes the binding problem by storing different types of information in one place and avoids the difficulty caused by their dispersal over separate stores.

Do these considerations imply that attempts to fractionate working memory should be abandoned and the unitary view proposed by Jones is correct? Not necessarily. Similar patterns of interference across modalities do suggest a common mechanism, but such a mechanism could supplement rather than replace modality-specific stores, as discussed in Section 2.2.2. Perhaps the strongest reason for not abandoning fractionation is that the irrelevant sound paradigm is a narrow basis on which to build a general theory. Thus a unitary account cannot begin to explain the array of evidence for dissociations from the wide variety of sources we have already discussed. Nevertheless, by suggesting an alternative interpretation of the irrelevant speech effect and drawing attention to the binding problem, the approach of Jones and his colleagues made an important contribution.

2.3.5 The episodic buffer

The three-component model of working memory failed to address a number of issues that can be related to the question of binding in one way or another. We have already touched on the added complexity needed to explain performance in working memory span tasks such as reading span (see Section 2.3.1). One aspect of the problem is that maintaining items whilst reading a series of sentences exceeds the storage capacity of the phonological loop. A related issue is the effect of expertise in boosting working memory capacity, as in the example of chess. In both cases performance benefits from chunking but the absence of any link between working memory and long-term

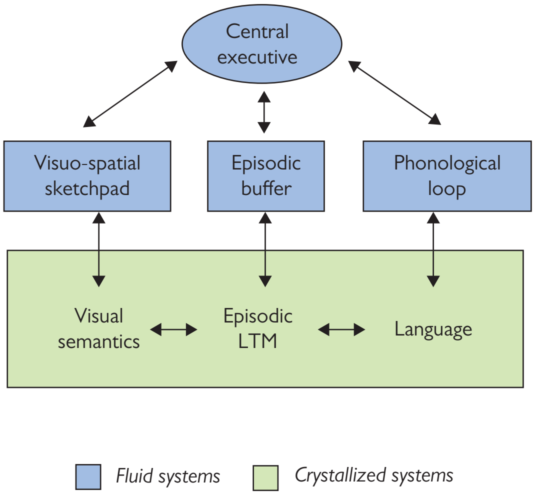


FIGURE 9.6 Revised structure of working memory.
Source: Baddeley, 2000

memory in early descriptions makes this difficult for the multi-component model to explain (see Figure 9.2). In an attempt to address such limitations, Baddeley (2000) proposed adding a fourth component, the **episodic buffer**, a limited capacity multi-dimensional store that backs up the central executive and at the same time provides a link with long-term memory (see Figure 9.6). The episodic buffer was assumed to bind information fed into it from peripheral short-term stores and from long-term memory to form multi-dimensional chunks, and to be able to hold a small number of such chunks at any given time. In agreement with Cowan (2005), the number of chunks was assumed to be in the region of four and to vary from one person to another, as reflected in individual differences in working memory span.

Subsequent research has focused on the question of whether the revised model is correct in its assumption that the processing power of the central executive is required to combine features into a single integrated representation or chunk. Baddeley and colleagues set out to explore this using the dual task methodology that had proved fruitful in developing the original concept of working memory. They were particularly interested in the prediction that blocking the central executive would interfere specifically with binding. In order to establish generality, memory for bindings was studied in two distinct contexts. One was the binding of visual features to form objects in visual working memory, and the other was the binding of words into chunks in memory for sentences.

In the visual domain, Luck and Vogel (1997) had shown previously that people can remember about four multi-featured objects, independently of the number of features in each, consistent with a chunk-based buffer. Wheeler and Treisman (2002) went on to show that increasing the attentional demands of the method used to probe retention affected memory for feature combinations (e.g. the colour and shape of an object) more than memory for individual features. (e.g. colour or shape). Wheeler and Treisman concluded that forming and maintaining feature bindings takes up attentional capacity. Allen *et al.* (2006) used an adaptation of Wheeler and Treisman's methodology in which participants performed a concurrent task during the encoding phase. They found that loading the central executive with a concurrent task such as counting backwards by threes impaired memory for feature bindings, but did so no more than memory for individual features. This pattern of equivalent disruption remained even when binding was made more difficult by separating the features of objects in space or time (Karlsen *et al.*, 2010) or presenting one feature visually and the other verbally (Allen *et al.*, 2009). However, memory for feature bindings was particularly sensitive to disruption from presenting further stimuli of the same type, even when participants were instructed to ignore them (Ueno *et al.*, 2011) and it is possible that a similar interference effect accounts for Wheeler and Treisman's (2002) findings. Overall, therefore, there is little evidence to support the hypothesis that the central executive is especially involved in forming and maintaining visual feature bindings.

In the verbal domain, chunking is reflected in the observation that memory span for sentences is considerably higher than standard word span, being of the order of 15 words (Brener, 1940). This 'sentence superiority effect' is attributed to the binding of words into meaningful chunks or phrases using information in long-term memory. In a series of dual task experiments, Baddeley *et al.* (2009) explored whether blocking the central executive would disrupt the binding of verbal information into chunks and therefore reduce the sentence superiority effect. In parallel with their results on visual feature binding, loading the central executive with an attention-demanding concurrent task led to an overall impairment in memory performance, but this was no greater for sentences than for scrambled word lists.

Baddeley *et al.* (2011) interpreted these unexpected findings as suggesting that binding is largely automatic, whether it be the chunking together of words using information in long-term memory or the integration of perceptual features into representations of objects. This in turn suggests a modification of the multi-component model in which the episodic buffer is a passive rather than an active system. The buffer is still assumed to serve the crucial integrative role of holding information from different sources bound into unitized episodes or chunks. However, the evidence suggests that binding itself takes place outside the episodic buffer and that the phonological loop and visuo-spatial sketchpad feed the episodic buffer directly rather than, as had been initially assumed (Baddeley, 2000), via the central executive (cf. Figure 9.6).

Baddeley *et al.* (2011) were keen to emphasize that they regard the multi-component approach to working memory as providing a broad framework that attempts to capture what we already know in a relatively simple way and generate further questions. The episodic buffer provides a neat illustration of how asking further questions led to a revision of the framework. Time will tell whether it can continue to grow incrementally or will ultimately have to be abandoned in favour of alternative approaches.

SUMMARY OF SECTION 2

- According to Baddeley and Hitch (1974), working memory is a multi-component system. The main components initially were the phonological loop, the visuo-spatial sketchpad, and the central executive.
- The phonological loop has been further fractionated into the phonological store and a control process of sub-vocal rehearsal. The visuo-spatial sketchpad has been divided into separate visual and spatial components. A possible kinaesthetic component of working memory has yet to be firmly established.
- The central executive may also involve several sub-components. However, these have yet to be clearly defined and understood.
- The operation of a fractionated system needs to ensure that multiple sub-systems do not result in incoherence. One aspect of this is the binding problem, which refers to the need for the system to keep track of information spread between different sub-systems.
- Concepts such as the episodic record (Jones, 1993) and the episodic buffer (Baddeley, 2000) may be viewed as different attempts to solve the binding problem.

3 VOCABULARY ACQUISITION

So far we have mentioned some but by no means all of the many functions of working memory and its subsystems. One that has been studied particularly closely is the role of the phonological loop in learning new vocabulary. The ability to store the sequence of phonemes making up a word is particularly important when encountering the word for the first time and retaining its spoken form long enough to learn it. The evidence comes from a variety of sources that include neuropsychological impairment, studies of individual differences in vocabulary size, and experimental studies of word-learning (see Baddeley *et al.*, 1998 for a review).

3.1 Neuropsychological evidence

Some of the clearest evidence linking the phonological loop with vocabulary acquisition comes from the previously described patient PV, whose phonological store had a markedly reduced capacity (see Section 2.2.3). PV had normal long-term memory for familiar items, but encountered profound difficulty in learning new word forms. Baddeley *et al.* (1988) showed this experimentally by testing her ability to learn pairings such as *Rosa–Svieti*, where the first word was in her native Italian and the second was an unfamiliar Russian word. The result was dramatic: PV showed no learning at all. However, when the second words were also Italian she showed normal learning. These observations establish a relationship between short-term phonological storage and long-term phonological learning. This in turn resurrects a classic debate about the relationship between short- and long-term memory. Similar patients, such as KF (see Section 1.2), who had impaired STM but normal LTM, were historically important to the argument for separate memory stores. That dissociation still stands, but the fact that PV could only learn pairings of familiar items (whose phonetic structure is already stored in LTM) indicates a more complex relationship between STM and LTM. This is illustrated in the partitioning of LTM into linguistic and episodic regions in Baddeley's (2000) revised model of working memory and the assumption that each region has a separate link to the phonological store (see Figure 9.6).

3.2 Individual differences

If learning new vocabulary items involves the capacity to hold a phonological sequence over a short interval, then the two abilities should be correlated with each other. A number of studies have shown that children's auditory digit span does indeed correlate with their performance on tests of vocabulary (see Baddeley *et al.*, 1998). Further evidence has come from studying the child's ability to repeat a non-word they have just heard (e.g. *Blonterstaping*). Non-word repetition was devised as a more demanding test of memory for phonological form than digit span, and is typically more highly correlated with vocabulary scores than digit span. Furthermore, individual differences in the capacity of the phonological loop predict children's performance on a simulated vocabulary learning task (Gathercole *et al.*, 1997). Understanding the causal basis of these relationships is far from straightforward, however. For example, longitudinal studies suggest that short-term phonological storage capacity limits vocabulary growth early on in child development, whereas later on vocabulary knowledge influences performance in phonological short-term memory tasks (Gathercole and Baddeley, 1989; see also Metsala, 1999).

As a postscript, it is interesting to note that measures of the phonological loop also correlate with vocabulary in second-language learning. For example, Service (1992) found that Finnish children's ability to repeat English-sounding non-words before starting to learn English predicted their English vocabulary some two years later. Furthermore, Papagno and Vallar (1995) showed that polyglots selected for being fluent in at least three languages had superior auditory digit span and non-word repetition when compared with controls. The polyglots were especially good at learning word–non-word pairs but were no better than controls at learning word–word pairs.

3.3 Experimental studies

Yet another way of assessing the involvement of the phonological loop in new word learning is to take an experimental approach. In a series of studies, Papagno and her colleagues investigated adults learning sets of either word–non-word pairs or word–word pairs. Papagno and Vallar (1992) showed that increasing the phonemic similarity of the non-words or their number of syllables impaired learning. However, corresponding manipulations in the word–word learning task had no effect. Furthermore, Papagno *et al.* (1991) found that articulatory suppression impeded the learning of word–non-word pairs but had no effect on learning word–word pairs. The absence of effects on the word–word learning task provides confirmation that the role of the loop is specific to learning novel words. These experimental differences between word–word and word–non-word learning fit well with the data on individual differences and patient PV in these same tasks.

SUMMARY OF SECTION 3

- The phonological loop is involved in learning new word forms but not new associations between familiar words. These tasks dissociate in the neuropsychological patient PV and healthy adults.
- Individual differences in vocabulary correlate with performance in tasks that reflect the capacity of the phonological loop in children and adults.
- However, the causal nature of the relationship between the phonological loop and vocabulary during development may be complex.

4 MODELLING THE PHONOLOGICAL LOOP

A number of attempts have been made to develop mathematical and computational models of the phonological loop (see for example Brown *et al.*, 2000; Burgess and Hitch, 1992, 1999, 2006; Page and Norris, 1998). Part of the impetus behind these efforts is the need to explain phenomena that the initial concept fails to address. For example, the phonological loop was originally presented as an account of immediate recall, and although Baddeley (2000) subsequently showed it as linked to long-term memory (see Figure 9.6), the revised model does not say how this link operates. Clearly, extra assumptions are needed to account for how phonological forms of newly learned words are acquired. Even within immediate recall, the phonological loop is far from providing a complete account. The most common type of error in the immediate recall of sequences of familiar items such as digits or letters is where the items themselves are correct but their order is confused. However, the original concept of the phonological loop does not explain how information about order is encoded or how order errors are generated. These omissions make a case for extending the concept of the phonological loop to account for long-term learning and serial ordering, while at the same time attempting to preserve its original insights.

One argument for using computational modelling to develop and express theories is the increasing complexity of the facts we have to explain. As should be evident from the present discussion, one strength of the original concept of the phonological loop is its simplicity and the ease with which it can be used to generate testable predictions. In passing we may also note that this same strength has also allowed investigators to assess whether each of its assumptions is correct (see Section 2.2.4). This is an important part of the normal scientific process. However, revising and extending the concept of the phonological loop to cope with errors and omissions runs the risk of generating an increasingly unwieldy informal account. In particular, adding capabilities for serial ordering and learning may render the concept too complex to work with. Moving from an informal, verbal–conceptual level of theorizing to a more explicit, computational model that can simulate human behaviour is one way of attempting to overcome this problem.

The most basic test of the adequacy of a computational model is whether it reproduces the human behaviour when presented with the same tasks. However, this is not necessarily a very convincing test, as the model-builder knows in advance the phenomena of interest and in general will have made sure the model succeeds in reproducing them. A more powerful test is to run further simulations in which the model is presented with novel experiments. The model's pattern of behaviour corresponds to its prediction about human behaviour in the same circumstances. The experiments can then be run with human participants to see whether the model's predictions are upheld (see for example Burgess and Hitch, 2006; Hitch *et al.*, 2009). Unfortunately, it is not quite as simple as this sounds, and there are many reasons for being cautious before embarking on computational modelling. One is that developing a mechanistic account involves making a large number of assumptions in order for the model to 'run'. Sometimes the challenge of justifying these assumptions is hard to meet. In the case of auditory–verbal STM we are fortunate that there is a wealth of published data with which to constrain model-building. However, the same cannot be said at present for executive function, and detailed computational modelling would almost certainly be premature in this case. In the following section, we describe briefly some constraints that influence the solution to the problem of how to handle serial order in the context of a computational model of the phonological loop. Note that we do not discuss models in detail, nor evaluate fully their ability to explain existing data. These are important topics, but would take us beyond our present goal of providing a basic introduction to the way modelling has begun to influence theorizing about working memory.

4.1 Serial order

The general problem of explaining serial order in behaviour is widely recognized and several types of mechanism have been proposed (Lashley, 1951). We will briefly describe some of these, bearing in mind that what interests us here is the specific question of what type of ordering mechanism underpins the operation of the phonological loop.

According to the chaining hypothesis, serial order is coded by forming associations between consecutive items (e.g. Jones, 1993; Wickelgren, 1965). However, although chaining might seem highly plausible, it encounters some basic problems. One is explaining recall of a sequence containing repeated items, such as the number *2835867*. If order is encoded as a chain of associations, then the repeated item (*8*) will be associated with not one but two following items (*3* and *6*). Consequently sequences containing repeated items should be difficult to recall and errors should occur after each occurrence of the repeated item. Although sequences containing repeats are more difficult to recall in verbal STM tasks, errors tend to occur *on* rather than *after* the repeated item (Jahnke, 1969). Further evidence against chaining comes from errors in recalling sequences in which phonemically similar items alternate with phonemically dissimilar items, such as *BXDJTQVR*. These errors show a characteristic zig-zag pattern as one goes through the list, with more errors on the similar-sounding items (i.e. *BDTV*) and fewer on the dissimilar items (Baddeley, 1968; Henson *et al.*, 1996). Furthermore, dissimilar items are recalled with the same accuracy in alternating lists as in pure lists where all the items are dissimilar (see Figure 9.7). According to chaining theories, extra errors ought to occur on the dissimilar items, as these follow similar cues. For these reasons, chaining seems unlikely to explain how the phonological loop deals with serial order. It is, however, a very pervasive concept and some models combine chaining with further assumptions and manage to achieve some degree of success (Botvinick and Plaut, 2006; Lewandowsky and Murdock, 1989).

One alternative to chaining is the positional hypothesis, according to which order is coded by associations between each item and a representation of its position within the sequence. In the simplest example of this type of model, Conrad (1965) assumed that verbal STM is composed of an ordered array of slots, each containing a successive item in a list. To remember the sequence, the contents of the slots are simply read out. This simple model has no problem explaining how sequences containing repeated items are

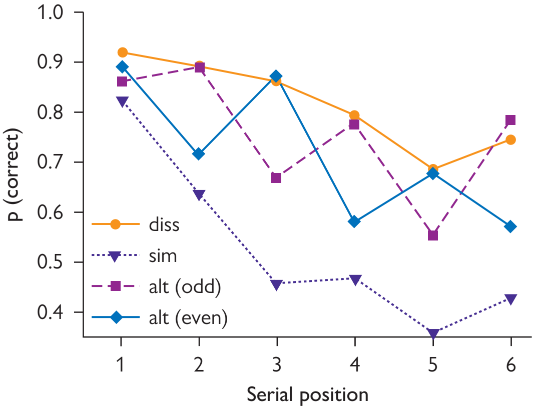


FIGURE 9.7 Serial position curves for the recall of six-item lists that varied in their phonemic composition. Diss: all items phonemically dissimilar; sim: all items phonemically similar; alt(odd)/alt(even): alternating phonemically similar and dissimilar items, with similar items in odd or even-numbered serial positions. *Source: Baddeley, 1968, Experiment V*

recalled. However, it cannot account for typical order errors in serial recall, where a common failure consists of an exchange between two adjacent items (e.g. recalling the sequence *318476205* as *318746205*; this is known as a transposition error). The probability of making a transposition error decreases with distance from the correct position (Healy, 1974). Estes (1972) proposed a mathematical model to account for this distribution of order errors, according to which positional information is encoded for each item and becomes less precise as a function of forgetting. In a related approach, some computational models (e.g. Burgess and Hitch, 1999, 2006; Brown *et al.*, 2000) assume order is coded by associations between items and a timing signal that varies with their position in a sequence. The timing signal provides an approximate coding of position and is used to explain the distribution of order errors in a somewhat similar way to Estes. One success of this approach is that it can explain the zig-zag variation of recall with position for lists of alternating phonemically similar and dissimilar items (e.g. *BXDJTQVR*). The Burgess and Hitch model (1999) achieves this by assuming that recall of each item is a two-stage process involving first using positional information to select a candidate item, and second retrieving the phonemic content of the selected item. Phonemic similarity of items is assumed to make the second of these two stages less efficient, but has no effect on the first stage. Burgess and Hitch (1999) report simulations by their model. For alternating conditions, these have the same zig-zag form as the experimental data (even though the simulations do not give enough 'primacy', i.e. decline in recall from the start of the list). Unsurprisingly, these zig-zag serial position curves and their interpretation have become an important focus of interest within the field. This has led to new findings and interpretations that go beyond the simple picture painted here (see for example Botvinick and Plaut, 2006; Farrell and Lewandowsky, 2003; Lewandowsky and Farrell, 2008).

A third point of view comes from non-associative models, according to which encoding the order of a sequence does not involve forming novel associations. For example, in the primacy model of Page and Norris (1998), differences in the activation levels of items in memory are used to encode information about their order. Page and Norris (1998) assume that each successive item in a sequence is encoded with a lower level of activation than its predecessor. This process forms a 'primacy gradient' of activation levels over the list. Recall of the items in the correct serial order involves an iterative process of choosing the most strongly activated item, then the next and so on. Although this model is radically different from positional accounts, it passes the test of being able to simulate zig-zag patterns in the recall of lists of phonemically similar and dissimilar items. It is interesting to note that the model achieves this by assuming that the primacy gradient is used to select each item for recall but that a second phonological stage is used to retrieve the item's phonemic composition. Thus the primacy model and the positional model of Burgess and Hitch (1999) share the idea of two stages in recall, the first processing order information and the second item information, but differ in how they assume the first stage in particular works.

So far, then, we can see that data on alternating lists is useful for ruling out chaining models but does not discriminate between positional and non-associative models. Fortunately, there are further data that help discriminate between these two classes of model. These relate to the temporal grouping effect, whereby presenting a sequence of items in rhythmic temporal groups brings about a marked reduction in order errors in immediate recall (Ryan, 1969). Thus, recall of a sequence such as *318476205* is more accurate if it is presented as groups of three items, i.e. *318,476,205* (where the commas denote pauses). Moreover, grouping changes the pattern of order errors. Instead of the most common mistake being to transpose an item to an adjacent position, errors of recalling an item in a corresponding position in a different group become much more frequent, as in *316,478,205*. These effects of grouping suggest a positional coding system in which position can be encoded at different levels. That is, a higher level codes the position of groups within a list and a lower level codes the position of items within groups. Hitch *et al.* (1996) show how their computational model captures these hierarchical effects of position in memory. Insofar as Page and Norris's (1998) primacy model encodes order on a single dimension, it cannot explain evidence for the coding of order on different levels. However, it would not be impossible to extend the model to include combinations of primacy gradients at different levels.

Modelling is presently developing in a number of broad directions. One is extending positional models of ordering at the lexical level to account for serial ordering at phonological and syllabic levels (Gupta and MacWhinney, 1997; Hartley and Houghton, 1996). This is important as language perception and production depend on simultaneous ordering at a number of different levels and understanding how these are integrated presents a serious challenge. Another direction is extending such models to account for long-term sequence learning (Burgess and Hitch, 2006; Hitch *et al.*, 2009). Again, this is important not only as part of the theoretical challenge of characterizing the relationship between working memory and LTM, but also as a pathway towards greater understanding of how new word-forms are learned in vocabulary acquisition.

SUMMARY OF SECTION 4

- One of the arguments for modelling is to go beyond the verbally expressed original concept of the phonological loop and address a wider range of phenomena such as serial ordering and non-word learning.
- Modelling is appropriate when we have a reasonably good general understanding of the system we are trying to model and there are extensive data with which to constrain model-building.
- The modelling of serial order in the phonological loop illustrates how existing data can be used to help make decisions about underlying mechanisms.
- The chaining hypothesis has difficulty explaining certain aspects of serial order recall. However, other hypotheses have had more success, for example, the positional hypothesis and non-associative accounts.

5 CONCLUSION

We have examined the concept of working memory as a multi-component system, with particular emphasis on phonological working memory and executive control. Taking a historical approach, we have traced the emergence of the concepts of the phonological loop, the visuo-spatial sketchpad, and the central executive, and the subsequent development of the model. Progress in tackling different aspects of working memory has developed at different rates. In the case of the phonological loop, experimental evidence from a variety of sources converged on a relatively simple two-component account. In turn, this account led to insights into a range of phenomena, including the development of STM, its neuropsychological impairment, and children's learning of vocabulary. Although the original model of the phonological loop has been shown to be inadequate in a number of details, it nevertheless preserves sufficient insights to have encouraged computational modellers to develop more detailed accounts that explain a greater range of phenomena.

In the case of the central executive the story is rather different. Here progress has been much slower and has consisted of various attempts to get an adequate conceptual handle on the problem. In this context the difficulty of devising satisfactory and reasonably well-understood tests of the various aspects of executive function that have been proposed should perhaps not surprise us. What is needed is a greater conceptual understanding of the various functions of executive control, one that goes beyond the promising beginning made by Norman and Shallice (1986) and develops the sorts of ideas discussed by Baddeley (1996).

In closing, we note that specifying the structure of working memory is useful but cannot be the whole story. A number of important issues have emerged from viewing working memory as a multi-component system. One of these concerns how such a system functions in a coherent and coordinated way, with the episodic buffer being proposed to address the particular issue of binding. A second issue concerns how working memory is linked to the cognitive system as a whole. Here too, work has begun in addressing the role of working memory in long-term learning and the effects of previously acquired knowledge on working memory. Only when broader issues such as these are addressed can we start to give a more precise and coherent account of the role of working memory in such ordinary but complex everyday activities as planning a shopping trip or reading a newspaper.

FURTHER READING

Andrade, J. (ed.) (2001) *Working Memory in Perspective*, Hove, Psychology Press. Chapters by experienced researchers present a critical assessment of the Baddeley and Hitch (1974) model of working memory.

Baddeley, A.D. (2000) 'Is working memory still working?', *American Psychologist*, vol.56, pp.851–64. In 2001 Alan Baddeley received the American Psychological Association's Award for Distinguished Scientific Contributions. This article presents his award address, which took the form of a personal review of the current state of the Baddeley and Hitch (1974) model of working memory.

Miyake, A. and Shah, P. (1999) *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, Cambridge, Cambridge University Press. Proponents of competing theoretical approaches to working memory were invited to present their views in a format that was designed to help clarify areas of agreement and disagreement.

Baddeley, A.D. (2007) *Working Memory, Thought, and Action*, Oxford, Oxford University Press. This is a broad-ranging book that reviews the current status of the multi-component model and extends it to topics such as emotion, consciousness, and neuroimaging.

REFERENCES

Allen, R.J., Baddeley, A.D., and Hitch, G.J. (2006) 'Is the binding of visual features in working memory resource-demanding?' *Journal of Experimental Psychology: General*, vol.135, pp.298–313.

Allen, R.J., Hitch, G.J., and Baddeley, A.D. (2009) 'Cross-modal binding and working memory', *Visual Cognition*, vol.17, pp.83–102.

Andrade, J. (2001) 'The working memory model: consensus, controversy, and future directions', in Andrade, J. (ed.) *Working Memory in Perspective*, Hove, Psychology Press.

Atkinson, R.M. and Shiffrin, R.M. (1971) 'The control of short-term memory', *Scientific American*, vol.225, pp.82–90.

Baddeley, A.D. (1966a) 'Short-term memory for word sequences as a function of acoustic, semantic and formal similarity', *Quarterly Journal of Experimental Psychology*, vol.18, pp.362–5.

Baddeley, A.D. (1966b) 'The influence of acoustic and semantic similarity on long-term memory for word sequences', *Quarterly Journal of Experimental Psychology*, vol.18, pp.302–9.

Baddeley, A.D. (1968) 'How does acoustic similarity influence short-term memory?' *Quarterly Journal of Experimental Psychology*, vol.20, pp.249–64.

Baddeley, A.D. (1983) 'Working memory', *Philosophical Transactions of the Royal Society of London B*, vol.302, pp.311–24.

Baddeley, A.D (1986) *Working Memory*, Oxford, Oxford University Press.

Baddeley, A.D. (1996) 'Exploring the central executive', *Quarterly Journal of Experimental Psychology*, vol.49A, pp.5–28.

Baddeley, A.D. (2000) 'The episodic buffer: a new component of working memory?' *Trends in Cognitive Sciences*, vol.4, pp.417–23.

Baddeley, A.D. (2007) *Working Memory, Thought, and Action*, Oxford, Oxford University Press.

Baddeley, A.D. and Hitch, G.J. (1974) 'Working memory', in Bower, G. (ed.) *The Psychology of Learning and Motivation: Advances in Research and Theory*, vol.8, New York, Academic Press.

Baddeley, A.D. and Larsen, J.D. (2003) 'The disruption of STM: a response to our commentators', *Quarterly Journal of Experimental Psychology*, vol.56A, pp.1301–6.

Baddeley, A.D. and Lieberman, K. (1980) 'Spatial working memory', in Nickerson, R.S. (ed.) *Attention and Performance VIII*, Hillsdale, NJ, Erlbaum.

Baddeley, A., Allen, R.J., and Hitch, G.J. (2011) 'Binding in visual working memory: the role of the episodic buffer', *Neuropsychologia*, vol.49, pp.1393–400.

Baddeley, A., Gathercole, S., and Papagno, C. (1998) 'The phonological loop as a language learning device', *Psychological Review*, vol.105, pp.158–73.

Baddeley, A.D., Hitch, G.J., and Allen, R.J. (2009) 'Working memory and binding in sentence recall', *Journal of Memory and Language*, vol.61, pp.438–56.

Baddeley, A.D., Lewis, V.J., and Vallar, G. (1984) 'Exploring the articulatory loop', *Quarterly Journal of Experimental Psychology*, vol.36A, pp.233–52.

Baddeley, A.D., Papagno, C., and Vallar, G. (1988) 'When long-term learning depends on short-term storage', *Journal of Memory and Language*, vol.27, pp.586–96.

Baddeley, A.D., Thomson, N., and Buchanan, M. (1975) 'Word length and the structure of short-term memory', *Journal of Verbal Learning and Verbal Behavior*, vol.14, pp.575–89.

Barrouillet, P. and Camos, V. (2001) 'Developmental increase in working memory span: resource sharing or temporal decay?' *Journal of Memory and Language*, vol.45, pp.1–20.

Beaman, C.P. and Jones, D. (1997) 'The role of serial order in the irrelevant speech effect: tests of the changing-state hypothesis', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.23, pp.459–71.

Botvinick, M. and Plaut, D.C. (2006) 'Short-term memory for serial order: a recurrent neural network model', *Psychological Review*, vol.113, pp.201–33.

Brener, R. (1940) 'An experimental investigation of memory span', *Journal of Experimental Psychology*, vol.26, pp.467–83.

Brown, G.D. and Hulme, C. (1995) 'Modelling item length effects in memory span: no rehearsal needed?' *Journal of Memory and Language*, vol.34, pp.594–621.

Brown, G.D., Preece, T., and Hulme, C. (2000) 'Oscillator-based memory for serial order', *Psychological Review*, vol.107, pp.127–81.

Brown, J. (1958) 'Some tests of the decay theory of immediate memory', *Quarterly Journal of Experimental Psychology*, vol.10, pp.12–21.

Burgess, N. and Hitch, G.J. (1992) 'Toward a network model of the articulatory loop', *Journal of Memory and Language*, vol.31, pp.429–60.

Burgess, N. and Hitch, G.J. (1999) 'Memory for serial order: a network model of the phonological loop and its timing', *Psychological Review*, vol.106, pp.551–81.

Burgess, N. and Hitch, G.J. (2006) 'A revised model of short-term memory and long-term learning of verbal sequences', *Journal of Memory and Language*, vol.55, pp.627–52.

Caplan, D. and Waters, G.S. (1999) 'Verbal working memory and sentence comprehension', *Behavioural and Brain Sciences*, vol.22, pp.77–126.

Case, R.D., Kurland, D.M., and Goldberg, J. (1982) 'Operational efficiency and the growth of short-term memory span', *Journal of Experimental Child Psychology*, vol.33, pp.386–404.

Conrad, R. (1965) 'Order errors in immediate recall of sequences', *Journal of Verbal Learning and Verbal Behavior*, vol.4, pp.161–9.

Conway, A.R.A. and Engle, R.W. (1994) 'Working memory and retrieval: a resource-dependent inhibition model', *Journal of Experimental Psychology: General*, vol.123, pp.354–73.

Cowan, N. (1988) 'Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system', *Psychological Bulletin*, vol.96, pp.341–70.

Cowan, N. (2005) *Working Memory Capacity*, Hove, Psychology Press.

Cowan, N., Day, L., Saults, J.S., Keller, T.A., Johnson, T., and Flores, L. (1992) 'The role of verbal output time in the effects of word length on immediate memory', *Journal of Memory and Language*, vol.31, pp.1–17.

Craik, F.I.M. and Lockhart, R.S. (1972) 'Levels of processing: a framework for memory research', *Journal of Verbal Learning and Verbal Behavior*, vol.11, pp.671–84.

Daneman, M. and Carpenter, P.A. (1980) 'Individual differences in working memory and reading', *Journal of Verbal Learning and Verbal Behavior*, vol.19, pp.450–66.

de Beni, R., Palladino, P., Pazzaglia, F., and Cornoldi, C. (1998) 'Increases in intrusion errors and working memory deficit of poor comprehenders', *Quarterly Journal of Experimental Psychology*, vol.51A, pp.305–20.

Dempster, F.N. (1981) 'Memory span: sources of individual and developmental differences', *Psychological Bulletin*, vol.89, pp.63–100.

de Renzi, E. and Nichelli, P. (1975) 'Verbal and non-verbal short term memory impairment following hemisphere damage', *Cortex*, vol.11, pp.341–53.

Duncan, J. and Owen, A.M. (2000) 'Common regions of the frontal lobe recruited by diverse cognitive demands', *Trends in Neurosciences*, vol.23, pp.475–83.

Engle, R.W., Carullo, J.W., and Collins, K.W. (1991) 'Individual differences in working memory and comprehension: a test of four hypotheses', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.18, pp.972–92.

Engle, R.W., Kane, M.J., and Tuholski, S.W. (1999a) 'Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and the functions of the prefrontal cortex', in Miyake, A. and Shah, P. (eds.) *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, Cambridge, Cambridge University Press.

Engle, R.W., Tuholski, S.W., Laughlin, J.E., and Conway, A.R.A. (1999b) 'Working memory, short-term memory and general fluid intelligence: a latent variable approach', *Journal of Experimental Psychology: General*, vol.128, pp.309–31.

Ericsson, K.A. and Kintsch, W. (1995) 'Long-term working memory', *Psychological Review*, vol.102, pp.211–45.

Estes, W.K. (1972) 'An associative basis for coding and organization in memory', in Melton, A.W. and Martin, E. (eds.) *Coding Processes in Human Memory*, Washington, DC, Winston.

Farah, M.J. (1984) 'The neurological basis of visual imagery: a componential analysis', *Cognition*, vol.18, pp.245–72.

Farrell, S. and Lewandowsky, S. (2003) 'Dissimilar items benefit from phonological similarity in serial recall', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.29, pp.838–49.

Friedman, N.P. and Miyake, A. (2004) 'The relations among inhibition and interference control functions: a latent variable analysis', *Journal of Experimental Psychology: General*, vol.133, pp.101–35.

Gathercole, S.E. and Baddeley, A.D. (1989) 'Evaluation of the role of phonological STM in the development of vocabulary in children: a longitudinal study', *Journal of Memory and Language*, vol.28, pp.200–13.

Gathercole, S.E. and Hitch, G.J. (1993) 'Developmental changes in short-term memory: a revised working memory perspective', in Collins, A.F., Gathercole, S.E., Conway, M.A., and Morris, P.E. (eds.) *Theories of Memory*, Hove, Lawrence Erlbaum Associates.

Gathercole, S.E., Adams, A-M., and Hitch, G.J. (1994) 'Do young children rehearse? An individual differences analysis', *Memory and Cognition*, vol.22, pp.201–7.

Gathercole, S.E., Hitch, G.J., Service, E., and Martin, A.J. (1997) 'Phonological short-term memory and new word learning in children', *Developmental Psychology*, vol.33, pp.966–79.

Gupta, P. and MacWhinney, B. (1997) 'Vocabulary acquisition and verbal short-term memory: computational and neural bases', *Brain and Language*, vol.59, pp.267–333.

Hartley, T. and Houghton, G. (1996) 'A linguistically constrained model of short-term memory for nonwords', *Journal of Memory and Language*, vol.35, pp.1–31.

Healy, A.F. (1974) 'Separating item from order information in short-term memory', *Journal of Verbal Learning and Verbal Behavior*, vol.13, pp.644–55.

Henson, R. (2001) 'Neural working memory', in Andrade, J. (ed.) *Working Memory in Perspective*, Hove, Psychology Press.

Henson, R.N.A., Norris, D.G., Page, M.P.A., and Baddeley, A.D. (1996) 'Unchained memory: error patterns rule out chaining models of immediate serial recall', *Quarterly Journal of Experimental Psychology*, vol.49A, pp.80–115.

Hitch, G.J. (1978) 'The role of short-term working memory in mental arithmetic', *Cognitive Psychology*, vol.10, pp.302–23.

Hitch, G.J., Brandimonte, M.A., and Walker, P. (1995) 'Two types of representation in visual memory: evidence from the effects of stimulus contrast on image combination', *Memory and Cognition*, vol.23, pp.147–54.

Hitch, G.J., Burgess, N., Towse, J.N., and Culpin, V. (1996) 'Temporal grouping effects in immediate recall: a Working Memory analysis', *Quarterly Journal of Experimental Psychology*, vol.49A, pp.116–39.

Hitch, G.J., Flude, B., and Burgess, N. (2009) 'Slave to the rhythm: experimental tests of a model for verbal short-term memory and long-term sequence learning', *Journal of Memory and Language*, vol.61, pp.97–111.

Hitch, G.J., Halliday, M.S., Dodd, A., and Littler, J.E. (1989) 'Development of rehearsal in short-term memory: differences between pictorial and spoken stimuli', *British Journal of Developmental Psychology*, vol.7, pp.347–62.

Hitch, G.J., Halliday, M.S., Schaafstal, A., and Schraagen, J.M. (1988) 'Visual working memory in young children', *Memory and Cognition*, vol.16, pp.120–32.

Hitch, G.J., Towse, J.N., and Hutton, U. (2001) 'What limits children's working memory span? Theoretical accounts and applications for scholastic development', *Journal of Experimental Psychology: General*, vol.130, pp.184–98.

Hulme, C., Thomson, N., Muir, C., and Lawrence, A. (1984) 'Speech rate and the development of short-term memory span', *Journal of Experimental Child Psychology*, vol.38, pp.241–53.

Jahnke, J.C. (1969) 'The Ranschburg effect', *Psychological Review*, vol.76, pp.592–605.

Jones, D.M. (1993) 'Objects, streams and threads of auditory attention', in Baddeley, A. and Weiskrantz, L. (eds.) *Attention, Awareness and Control*, Oxford, Oxford University Press.

Jones, D.M., Farrand, P., Stuart, G., and Morris, N. (1995) 'Functional equivalence of verbal and spatial information in serial short-term memory', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.21, pp.1008–18.

Jones, D.M., Hughes, R.W., and Macken, W.J. (2006) 'Perceptual organization masquerading as phonological short-term storage: further support for a perceptual-gestural view of short-term memory', *Journal of Memory and Language*, vol.54, pp.265–81.

Just, M.A. and Carpenter, P.A. (1992) 'A capacity theory of comprehension: individual differences in working memory', *Psychological Review*, vol.99, pp.122–49.

Karlsen, P., Allen, R.J., Baddeley, A.D., and Hitch, G.J. (2010) 'Binding across space and time in visual working memory', *Memory & Cognition*, vol.38, pp.292–303.

Klauer, K.C. and Zhao, Z. (2004) 'Double dissociation in visual and spatial short-term memory', *Journal of Experimental Psychology: General*, vol.133, pp.355–81.

Lashley, K. (1951) 'The problem of serial order in behavior', in Jeffress, L.A. (ed.) *Cerebral Mechanisms in Behavior*, New York, Wiley.

Lewandowsky, S. and Farrell, S. (2008) 'Phonological similarity in serial recall: constraints on theories of memory', *Journal of Memory and Language*, vol.58, pp.429–48.

Lewandowsky, S. and Murdock, B.B. Jr. (1989) 'Memory for serial order', *Psychological Review*, vol.96, pp.25–57.

Lhermitte, F. (1983) '"Utilization behaviour" and its relation to lesions of the frontal lobes', *Brain*, vol.106, pp.237–55.

Logie, R.H. (1995) *Visuo-Spatial Working Memory*, Hove, Lawrence Erlbaum Associates.

Lovatt, P.J., Avons, S.E., and Masterson, J. (2000) 'The word-length effect and disyllabic words', *Quarterly Journal of Experimental Psychology*, vol.53A, pp.1–22.

Luck, S.J. and Vogel, E.K. (1997) 'The capacity of visual working memory for features and conjunctions', *Nature*, vol.390, pp.279–81.

Macken, W.J. and Jones, D.M. (1995) 'Functional characteristics of the "inner voice" and the "inner ear": single or double agency?' *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.21, pp.436–48.

Metsala, J.L. (1999) 'Young children's phonological awareness and nonword repetition as a function of vocabulary development', *Journal of Educational Psychology*, vol.91, pp.3–19.

Miller, G.A. (1956) 'The magical number seven plus or minus two: some limits on our capacity for processing information', *Psychological Review*, vol.63, pp.81–97.

Miyake, A. and Shah, P. (1999) *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, Cambridge, Cambridge University Press.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., and Howerter, A. (2000) 'The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis', *Cognitive Psychology*, vol.41, pp.49–100.

Murdock, B.B. Jr. (1967) 'Recent developments in short-term memory', *British Journal of Psychology*, vol.58, pp.421–33.

Nairne, J.S. (2002) 'Remembering over the short-term: the case against the standard model', *Annual Review of Psychology*, vol.53, pp.53–81.

Naveh-Benjamin, M. and Ayres, T.J. (1986) 'Digit span, reading rate, and linguistic relativity', *Quarterly Journal of Experimental Psychology*, vol.38A, pp.739–52.

Nicolson, R.S. (1981) 'The relationship between memory span and processing speed', in Friedman, M., Das, J.P., and O'Connor, N. (eds.) *Intelligence and Learning*, New York, Plenum Press.

Norman, D.A. and Shallice, T. (1986) 'Attention to action: willed and automatic control of behavior', in Davidson, R.J., Schwartz, G.E., and Shapiro, D.E. (eds.) *Consciousness and Self-Regulation: Advances in Research and Theory, vol.4*, New York, Plenum Press.

Page, M.P.A. and Norris, D.G. (1998) 'The primacy model: a new model of immediate serial recall', *Psychological Review*, vol.105, pp.761–81.

Page, M.P.A. and Norris, D.G. (2003) 'The irrelevant sound effect: what needs modeling and a tentative model', *Quarterly Journal of Experimental Psychology*, vol.56A, pp.1289–1300.

Palmer, S. (2000a) 'Working memory: a developmental study of phonological recoding', *Memory*, vol.8, pp.179–94.

Palmer, S. (2000b) 'Phonological recoding deficit in working memory of dyslexic teenagers', *Journal of Research in Reading*, vol.23, pp.28–40.

Papagno, C. and Vallar, G. (1992) 'Phonological short-term memory and the learning of novel words: the effect of phonological similarity and item length', *Quarterly Journal of Experimental Psychology*, vol.44A, pp.47–67.

Papagno, C. and Vallar, G. (1995) 'Short-term memory and vocabulary learning in polyglots', *Quarterly Journal of Experimental Psychology*, vol.48A, pp.98–107.

Papagno, C., Valentine, T., and Baddeley, A.D. (1991) 'Phonological short-term memory and the foreign-language vocabulary learning', *Journal of Memory and Language*, vol.30, pp.331–47.

Paulesu, E., Frith, C.D., and Frackowiack, R.S.J. (1993) 'The neural correlates of the verbal component of working memory', *Nature*, vol.362, pp.342–4.

Reason, J. (1984) 'Lapses of attention in everyday life', in Parasuraman, R., Davies, R., and Beatty, J. (eds.) *Varieties of Attention*, Orlando, FL, Academic Press.

Ryan, J. (1969) 'Grouping and short-term memory: different means and patterns of grouping', *Quarterly Journal of Experimental Psychology*, vol.21, pp.137–47.

Salamé, P. and Baddeley, A.D. (1982) 'Disruption of short-term memory by unattended speech: implications for structure of working memory', *Journal of Verbal Learning and Verbal Behavior*, vol.21, pp.150–84.

Service, E. (1992) 'Phonology, working memory, and foreign-language learning', *Quarterly Journal of Experimental Psychology*, vol.45A, pp.21–50.

Shallice, T. (2002) 'Fractionation of the supervisory system', in Stuss, D.T. and Knight, R.T. (eds.) *Principles of Frontal Lobe Function*, New York, Oxford University Press.

Shallice, T. and Warrington, E.K. (1970) 'Independent functioning of verbal memory stores: a neuropsychological study', *Quarterly Journal of Experimental Psychology*, vol.22I, pp.261–73.

Smyth, M.M. and Waller, A. (1998) 'Movement imagery in rock climbing: patterns of interference from visual, spatial and kinaesthetic secondary tasks', *Applied Cognitive Psychology*, vol.12, pp.145–57.

Towse, J.N. (1998) 'On random generation and the central executive of working memory', *British Journal of Psychology*, vol.89, pp.77–101.

Towse, J.N., Hitch, G.J., and Hutton, U. (1998) 'A re-evaluation of working memory capacity in children', *Journal of Memory and Language*, vol.39, pp.195–217.

Towse, J.N., Hitch, G.J., and Hutton, U. (2000) 'On the interpretation of working memory span in adults', *Memory and Cognition*, vol.28, pp.341–8.

Turner, M.L. and Engle, R.W. (1989) 'Is working memory capacity task dependent?' *Journal of Memory and Language*, vol.28, pp.127–54.

Ueno, T., Allen, R.J., Hitch, G.J., Baddeley, A.D., and Saito, S. (2011) 'Disruption of visual feature binding in working memory', *Memory and Cognition*, vol.39, pp.12–23.

Vallar, G. and Baddeley, A.D. (1984) 'Fractionation of working memory: neuropsychological evidence for a phonological short-term store', *Journal of Verbal Learning and Verbal Behavior*, vol.23, pp.151–61.

Wheeler, M.E. and Treisman, A.M. (2002) 'Binding in short-term visual memory', *Journal of Experimental Psychology: General*, vol.131, pp.48–64.

Wickelgren, W.A. (1965) 'Short-term memory for repeated and non-repeated items', *Quarterly Journal of Experimental Psychology*, vol.17, pp.14–25.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

This page intentionally left blank

PART 4

THINKING

Introduction

10 Problem solving
Alison J.K. Green and Ken Gilhooly

11 Judgement and decision making
Peter Ayton

12 Reasoning
Mike Oaksford

INTRODUCTION

In Part 4, the focus shifts to what have been termed thinking processes. Specifically, the chapters address three distinct kinds of thinking that arise in different kinds of task – in general terms, these are tasks that require us to solve problems (Chapter 10), come to judgements and make decisions (Chapter 11) and to reason and draw conclusions (Chapter 12).

In Chapter 10, Alison Green and Ken Gilhooly address human problem solving. You might think that problem solving is a somewhat artificial activity, inspired by abstract and contrived problems such as crossword puzzles or the Rubik's cube. But problem solving in cognitive psychology is intended to encompass a wide range of activities in which we need to identify the solution to a current problem. Everyday problems range from the easy, such as how to make a cup of tea in someone else's kitchen, to the complex, such as how to achieve career success. Everyday problems are not always easily defined – think of the different problems that need solving in order to achieve a successful career – and so psychologists have often relied on more formally specified problems, of which the authors of Chapter 10 provide numerous examples.

In Chapter 1 it was pointed out that cognitive psychology tends to avoid the study of individual differences because it aims to understand cognitive processes in general. However, in Chapter 9 of Part 3, we saw how the study of individual differences can help us to evaluate theories of cognition and, conversely, how the theories can help us to understand the nature of the individual differences. In Chapter 10, the importance of individual differences is emphasized again. Firstly, the authors note that some individuals are novices in solving certain classes of problem and some are expert. In playing chess, for example, some individuals achieve grand master status whereas others, while knowing the rules of the game, are considerably less skilled. The problem-solving approaches of these groups differ in important ways, and so it is not true to say that people in general tackle chess problems in the same way. Secondly, as the authors also point out, experts themselves differ from one another in relevant ways – that is, they do not form a homogeneous group – and novices also differ from one another.

Certain aspects of the study of problem solving that are raised in Chapter 10 become themes for the whole of Part 4. One is that cognitive psychologists place at least as much emphasis on the study of errors in problem solving as they do on cases of success. Indeed, as we shall see, errors provide important information concerning underlying cognitive processes, and this theme is continued in Chapters 11 and 12.

Another theme established in Chapter 10 is the importance of establishing a framework within which phenomena can be analysed and understood, and which in turn can be used to derive new research questions. In Chapter 10, Alison Green and Ken Gilhooly introduce the notion of a 'state–space' diagram and show how this notion can help us to understand problems and to analyse human performance when attempting to solve them.

In Chapter 11, Peter Ayton introduces the topics of judgement and decision making. How do we form judgements and make decisions, and how best should we understand and analyse these cognitive activities? One thing the author makes clear from the outset is that it is possible to develop different kinds of theory depending on the starting point and purpose of the theorist. A theory could, for example, emphasize how judgement and decision-making processes *ought* to proceed, and use this as a basis for analysing human performance. Alternatively, a theorist could take as their starting point an understanding of how people *actually* make decisions, including poor ones. That there are these two approaches – normative and descriptive – continues the theme, established in Chapter 10, of the importance of researchers establishing an appropriate explanatory framework. In essence, the chapter can be seen as an extended discussion of whether two particular normative approaches provide an adequate understanding of human judgement and decision making.

In the first part of Chapter 11, Peter Ayton discusses the normative theory of subjective expected utility and its use as a vehicle for understanding human decision making. As you will see, this theory requires us to express the likelihood of particular outcomes as mathematical probabilities, though you may be relieved to know that the key mathematical ideas are relatively simple. Probabilities are also required to understand a normative approach to judgement under uncertainty, which involves the application of Bayes' Theorem.

While the normative approaches rely on mathematical formulations of how judgement and decision making ought to proceed, the majority of the chapter discusses evidence that human performance actually departs from these mathematical standards – that human performance is characterized by apparent errors. The chapter compares such normative theories with descriptive accounts of actual human performance – prospect theory and the heuristics and biases approach. Although these accounts are seen ultimately to be more successful, the normative approaches nonetheless play an important role in helping researchers to develop a more appropriate explanatory framework. In particular, observations that human performance deviates systematically from mathematical standards have provided researchers with extremely valuable information.

The particular descriptive approaches discussed in the chapter are not without their problems, however. One important line of criticism comes from an approach that considers the adaptive function, in evolutionary terms, of decision-making processes. Some researchers have argued that such processes would have evolved to be 'fast and frugal', and the chapter cites evidence in favour of this view.

Some of these concerns can also be found in Chapter 12. In this chapter, Mike Oaksford shows how research into reasoning also started from a formal framework for understanding how people ought to reason – that of logic. The first part of Chapter 12 outlines the nature of logic, and in particular the forms of reasoning that are taken to be logically valid and those taken to be logically invalid. As with the normative approaches discussed in Chapter 11, logic provides a valuable framework for trying to understand human reasoning. It has also helped researchers to establish the core phenomena of reasoning, and to develop particular tasks to investigate these.

As with Chapter 11, a key observation also running throughout Chapter 12 is that human reasoning systematically departs from the normative standards established by logic. Predicting and understanding these logical errors has thus become an important benchmark for theories of human reasoning. A number of such theories have been developed, and the chapter focuses mostly on three approaches – mental logic, mental models, and the probabilistic approach. Each of these approaches is evaluated against evidence that has accrued from the use of two key tasks – conditional inference and the Wason selection task – and we see again the use of expected utilities and also Bayes' Theorem. It is worth noting that Chapter 12 uses the terms Type I and Type II processing in a completely different sense to the way in which they were used in Chapter 8, and the usages should not be confused.

Specifically in connection with a version of the Wason selection task, the chapter also discusses a fourth approach to reasoning. This approach emphasizes the importance of theories of reasoning positing processes that can be seen to have an evolutionarily adaptive function. Mike Oaksford also discusses the relationship between logical reasoning and IQ, showing once again the importance of individual differences.

As suggested above, there are a number of themes running throughout the three chapters of Part 4. One theme that has not been mentioned so far is rationality. Errors, or departures from a logical or mathematical standard, could be taken as signs of the intrinsic irrationality of human thought. After all, so the argument goes, if human beings are rational they ought to solve problems, make judgements and decisions, and reason according to certain standards, often assumed to be provided by formal models, such as mathematics and logic. Departures from these mathematical and logical standards would then be signs of irrationality. However, although these three chapters do not take rationality as their central focus, the establishment of theories of actual human performance provide grounds for understanding rationality differently. In connection with Chapter 11, for example, perhaps it would be rational to use heuristics and biases to come to a judgement, or, related to Chapter 12, perhaps it would be rational to rely on a probabilistic method for tackling a task, even if this sometimes generates logical 'errors'.

Finally, one thing you may notice about the three chapters in this part is that there is very little mention of computer modelling, neuropsychology, or neuroimaging. Although this inevitably reflects to a degree the practical limits imposed by writing a chapter to a certain length, it also reflects a particular emphasis common to the three chapters on trying to develop an appropriate explanatory framework. If you recall the discussion of Marr's levels of explanation in Chapter 1, it is as if cognitive psychologists studying thinking are still trying to establish what is computed when we reason, or make decisions, or solve problems. Consistent with this is the fact that some researchers are appealing to evolution to help provide an explanatory framework. The suggestion that researchers are still grappling with difficult questions at Marr's computational level provides one way of understanding the emphasis on formal approaches – such as state–space diagrams in Chapter 10, probability in Chapter 11, and logic in Chapter 12. Such formal approaches provide an idealized model of what needs to be computed, i.e. Marr's level 1 (idealized because we observe systematic departures from this in actual human thinking). The combination of these models and observations of systematic human 'error' provides researchers with an effective means for analysing human thinking. As research develops further, and detailed questions subsequently arise concerning the actual processes by which thinking is achieved, neuropsychological, neuroimaging, and computer modelling work is likely to become much more relevant.

CHAPTER 10

PROBLEM SOLVING

Alison J.K. Green and Ken Gilhooly

1 INTRODUCTION

Problem solving is an essential, familiar, and pervasive part of everyday life. Examples are all around us. Consider an infant trying to fit shapes into the appropriate holes of a shape-sorting toy, or a child trying to count out the correct sum of money to buy a new music CD, or perhaps an adult weighing up the pros and cons of a job offer. While we shall be looking here at examples of human problem solving, problem solving occurs in animal life too. Naturally occurring instances include tool use and searching for food. Problem solving in all its manifestations is an activity that structures everyday life in meaningful ways.

By studying the myriad ways in which we solve problems, we hope to learn how problems are solved effectively and to understand what goes wrong when they are not. Why should we be interested in finding out about *unsuccessful* problem solving? An interesting aspect of failure involves investigating the errors that people make in order to understand why a particular error occurred and to try to prevent it from happening again. Some errors are made with little cost (for example, sprinkling coffee instead of sugar over cornflakes at breakfast) but other errors can be quite catastrophic (for example, an oil-laden vessel running aground). Diagnosing errors and redesigning tasks to guard against critical errors are important applications for problem-solving research. We discuss others later on in the chapter.

Where does problem solving ‘sit’ in relation to other areas of psychology? Problem solving is an activity that draws together the various different components of cognition. For instance, linguistic skills are used to read about a political problem and engage in a debate about it, but language may also enhance our problem solving capabilities, as we shall see later. Visual perception is necessary for understanding a graphically-presented engineering problem and for drawing a solution. We use memory to recover any prior knowledge we might have that could be relevant to solving a new problem, and attention plays a role in all problem solving.

Problem solving takes place over time, interleaving a range of cognitive processes and drawing upon pieces of knowledge that are represented in various ways. The notion of 'representation' is central to cognitive psychology, as you will see later in Chapter 17. For now, we shall ask you to assume that information used in problem solving comes to be internally represented.

Because problem solving occurs over time, we need to study not just the cognitive processes and mental representations involved in problem solving, but also the ways in which these processes and representations interact with others. Problem solving then, like reasoning, judgement, and decision making, is an activity that necessarily draws upon a range of cognitive processes.

In fact, problem solving often involves reasoning, judgement, and decision making. For instance, a general practitioner gathering information about a sick patient's symptoms may deduce from the description the patient gives that the problem is a bacterial, rather than a viral, infection. The doctor may then make a judgement about the severity of the infection, before making a decision on an antibiotic to prescribe.

In this chapter, we examine the ways in which individuals approach a variety of problem types, ranging from simple puzzle problems to more complex real-world problems. Everyday problems can be complex and challenging, with constraints in operation that mean that the solution we choose or find may not be an ideal one. As individuals we can, if we are reasonably adept, persistent, or just plain lucky, solve many of the problems that come our way. Quite often, though, our initial attempts fail and we have to turn to another source to help solve the problem – a manual, for instance, in the case of a tricky computer installation problem, or perhaps someone knowledgeable in the problem area if all else fails.

While some problems may be viewed as unwelcome obstacles, to be avoided where possible, there are occasions where we keenly seek out problems to occupy our time. An expert mathematician, for instance, may spend hours identifying a problem, primarily for the pleasure derived in exploring and solving it. Similarly, make-over television programmes can be very entertaining as viewers watch an undecorated room or a derelict garden transformed by the experts in a matter of days into something quite different, or observe an antiques expert establishing what an item is and then estimating its worth. Some of us undoubtedly while away the hours working on tricky crossword puzzles or computer games, or trying to make (or repair) something at home.

Our aim in this chapter is to present an overview of research on problem solving. In doing this we have had to be selective, and have elected to present work that we believe has been both influential and interesting to try to give you a flavour of what has been going on in the field. As you read on, you will learn that how people represent problems is a principal determinant of problem-solving success. Much of the research we shall examine addresses the question of which factors influence the construction of a problem representation.

We aim to show you how ideas about problem solving have developed and changed. You will learn that early work on problem solving was often confined to puzzle problems, and that later on researchers became interested in more complex domains, where knowledge and experience are central to successful problem solving. The issues we shall explore centre on the nature of problem solving and the relationships between problem solving, learning, experience, and creativity. The kinds of questions we shall be asking include:

1. What are the different forms of problem-solving activity?
2. How do we solve different sorts of problems?
3. Why are representation and similarity important?

First, we shall try to define a 'problem' and then look at conceptions of problems and problem solving.

1.1 What is a 'problem'?

Before reading on, try Activity 10.1.

Consider the following examples of problems that a given individual might come across, some more commonplace than others:

1. Who can I ask to babysit the children so that I can go out next Thursday evening?

✱ ACTIVITY 10.1

What do you think are the defining attributes of problems? You will probably draw upon some examples from your own experience to help you. You might like to think back to Chapter 5 to help you think of problems in terms of properties, categories, and so on. Make a list of all the attributes you can think of, and then try to construct a sentence or two defining problems. Try not to spend more than a couple of minutes on this.

COMMENT

The answer to the question, 'What is a problem?' is not at all easy, as the exercise shows. You may find that you want to vary your definition depending upon the type of problem you have in mind. You may find that you cannot come up with a definition at all or that you came up with several and cannot choose between them. Of course, if it is difficult to define what we mean by a 'problem', then it becomes even more difficult to construct models and theories of problem-solving behaviour, and to compare and contrast such models and theories. Clear definitions are therefore important at the outset.

2. How can I make sure that the stone I have just played in my game of Go[1] 'lives'?
3. Is there a way I can arrange some paper pattern pieces on my dress material so that all the pieces fit and I don't have to buy any more material?

The problem in the first example is finding a babysitter, which could involve searching through an address book, recovering some names from memory, or calling round on a friend and asking for a favour. If these fail to produce a name then other options include carrying out a more extensive search. A bit of inspired guesswork might lead to an internet babysitting site and locating a babysitter to solve the problem. Notice that there are several ways to satisfactorily solve this problem, and that the possible solutions vary in degree of novelty. The availability of a possible solution method may well vary too, depending upon the context (is there time to explore different possible solutions to the problem?), social setting (is there a network of likely babysitters to call upon?), and culture (is it acceptable to use an internet babysitting agency?).

The problem in the second example centres on the ancient Korean game of Go. Here, the problem seems more to do with experience, knowledge, and skill, although motivation (does playing badly matter?), personality and emotion (are there personal costs in playing badly?), and cultural factors (different cultures have different conventions for Go) may well be involved too. Again, there are different solutions available, in that a number of different moves may achieve the goal of ensuring the 'survival' of the stone in question.

The final problem is different again, because it involves perception in 'seeing' how to lay all the pieces out, together with some creative or lateral thinking in optimizing layout so that all the pieces do indeed fit correctly. There may be one or more possible ways to arrange the pieces and solve the problem, one of which may be better (for example, in ensuring that cut pieces of fabric fit together in a way that matches up a pattern at seams).

These examples show that while problems do share some common characteristics (see the discussion of concepts in Chapter 5), it is also true that different problems are affected by different factors, both internal (for example, motivation and personality) and external (for example, social and cultural factors).

Duncker (1945, p.1) offered a concise definition of a problem that captures something of the essence of our everyday experience of problems. He wrote that: 'a problem exists when a living organism has a goal but does not know how this goal is to be reached'. The definition is still serviceable today because it conveys the notion of a 'gap' between a current state and a goal or desired state. If there are no obstacles preventing the individual from moving from the current state to the desired state, then there cannot be said to be a problem. Problems, then, consist of three components: a starting state, a goal state, and a set of available actions to move from the starting state to the goal state. According to this type of definition, what constitutes a problem for one individual may not be a problem for another. For instance, a moderate Go player might have some difficulty in ensuring that a newly placed stone survives in their current game if their opponent is a much stronger player. The stronger opponent, however, will almost certainly have considerably less difficulty in making sure their stones live.

---

[1] The aim in Go is to use stones (one player takes black stones, the other white) to surround territory on a board. A stone (or stones) 'lives' if it cannot be surrounded and therefore removed from the board. Territory is 'won' if stones of one colour completely surround stones of the other colour, and the winner is the player who surrounds the most territory.

So far, we have tried to present some defining characteristics of problems. Before we move on to discuss research on problem solving, we want to draw your attention to one of the principal methods used in problem-solving research: protocol analysis.

1.2 Protocol analysis in problem-solving research

Cognitive scientists make extensive use of a method known as 'protocol analysis'. Protocol analysis is one of a cluster of methods for analysing verbal data. At the core of the approach is the view that information represented in working memory may be verbalized without the sequence of thoughts being disrupted. Verbalization of such heeded information may take place either directly if that information is in verbal form, or through transformation if in non-verbal form. Heeded information then (the contents of working memory at a given point) may reflect both recently encoded information and information retrieved from long-term memory. Information retained in long-term memory must first be transferred to working memory before it can be reported. Thus, the 'protocol' of protocol analysis is a verbal account of information that is heeded as a task is carried out. The protocol that results from thinking aloud is assumed to preserve the order in which information has been heeded. Using careful instructions, and with a little practice, most people can 'think aloud', either while working on a task, or immediately after completing a task.

Protocol analysis depends upon fundamental assumptions, the most basic of which are that cognition is information processing, that information is stored in different memory stores, and that recently acquired information is retained in working memory. Protocol analysis then centres on elucidating thought processes. In contrast, 'verbal analysis' (Chi, 1997) focuses on the content of verbal data, and thus the emphasis is on the representation of knowledge that is used to support problem solving rather than the underlying processes. A further key distinction between the two methods centres on how the data are collected. Protocol analysis aims to elucidate what is heeded as a task is carried out, while verbal analysis is concerned with the explanations, rationalizations, and justifications people produce. Instructions to think aloud in protocol analysis are therefore deliberately geared towards not altering the individual's normal approach to performing the task, whereas the procedure used to generate data for verbal analysis will often explicitly invite the individual to generate self-explanations or rationalizations, activities that are in fact likely to alter performance of the task. A good example of this is the study of the role of self-explanation in learning by Chi *et al.* (1989), which we describe later in Section 5.2.1.

Protocol analysis has many uses, particularly in helping to identify differences between individuals in terms of information heeded, and processes and strategies used, as a task is carried out. Let us suppose that our research question centres on investigating cognitive processes in arithmetic, and that we have asked two individuals to think aloud while calculating the sum of 63 + 37. Both give the answer '100'. Did both arrive at the answer in the same way? One way of addressing this question is to compare the verbal reports produced:

> *First individual: 'OK, what is the sum of 63 plus 37? Easy – that's 100'. Second individual: 'What is the sum of 63 plus 37? 60 plus 30 . . . 60, 70, 80, 90. 3 plus 7 is 10. 90 plus 10 is 100. It's 100.'*

The first individual simply reads out the problem statement and then reports the answer. There is little evidence of any problem solving here, and the answer appears to be readily available – it is as if the individual is retrieving a number fact. The second individual also begins by reading the problem statement, but then goes about the problem rather differently. The protocol suggests that their strategy is a 'counting on' strategy, starting with 60, then counting on 30, giving 90. They then add the units 3 and 7, giving 10, and finally add 10 to 90 to give the answer. The example shows that different people can arrive at the same answer, but use different methods. It also shows that protocol analysis can reveal useful information about strategies underlying behaviour.

Protocol analysis is a very useful tool for identifying different strategies that people use in problem solving – strategies that may not be obvious from problem solutions alone. (In Chapter 19 you will encounter some models of the ways in which simple arithmetic problems, like those above, may be solved.) Of course, there are situations where protocol analysis is not a suitable approach (for example, where the requirement to think aloud might actually change the way in which the task is carried out).

It is important to recognize that thinking aloud is not the direct externalization of our cognitive processes. Rather, mental processes may be inferred through the careful analysis of verbal protocols. We illustrate an application of protocol analysis in Box 10.1.

BOX 10.1 METHODS Protocol analysis applied to medical diagnosis

Medical diagnosis is a complex skill, requiring the clinician to bring to bear his or her knowledge and skill in accurately diagnosing a given patient's disorder. Expert clinicians have acquired both biomedical and clinical knowledge. Biomedical knowledge includes knowledge of anatomy, biochemistry, and physiology, while clinical knowledge is often expressed in terms of associations between symptoms or clinical findings and disease categories. There has been some debate over the extent to which expert clinicians use biomedical knowledge in making diagnoses. Lesgold *et al.* (1998) found that expert clinicians made extensive use of biomedical knowledge, whereas Boshuizen and Schmidt (1992) found that they made very little use of it.

Gilhooly *et al.* (1997) hypothesized that when experts can use contextual information (e.g. patient's age, gender, and lifestyle habits) to aid a diagnosis, their use of biomedical knowledge may be suppressed. Gilhooly *et al.* tested this hypothesis through their analysis of think-aloud protocols produced by clinicians varying in skill level. They asked a group of clinicians to interpret electrocardiogram (ECG) trace information. The ECG is regularly used to assess the electrical activity of the heart, and to help identify abnormal patterns of activity that might indicate an underlying problem. Skill is required to interpret an ECG trace and to use this in making an accurate diagnosis, which then becomes the basis for a patient's treatment regime.

Gilhooly *et al.* asked groups of registrars (the 'experts'), house officers (the 'intermediates'), and third-year medical students (the 'novices') to think aloud while they studied and diagnosed eight different ECG traces, presented with no context information. They then analysed the protocols, examining them for evidence of biomedical and clinical knowledge. For example, use of key terms such as 'polarization', 'activation', or 'conducting' was categorized as biomedical reference. Use of words such as 'chronic' or 'hypertension' was classified as clinical reference. Clinicians also described the ECG traces directly in their protocols, giving a third category of words. In this way, the protocols produced by the clinicians were segmented into much smaller chunks, corresponding to clinical or biomedical inferences or trace descriptions.

Reassuringly, the more experienced and skilled the clinicians were, the more accurate their diagnoses. The results of the protocol analysis showed that more skilled clinicians made more extensive use of their biomedical knowledge than less skilled clinicians, particularly in evaluating possible diagnoses. They also made more use of their clinical knowledge than the less skilled clinicians.

What does this study tell us? First, it resolves the apparently discrepant findings in the literature. Increased use of biomedical knowledge *is* associated with expertise, when clinicians are not able to use shortcuts to aid a diagnosis. Second, the study shows that protocol analysis can be a very useful tool in helping us to understand problem solving in real-world situations. Verbal protocols can give valuable insights into knowledge and processes involved in problem solving.

SUMMARY OF SECTION 1

- Problems involve a start state, a goal state, and a set of actions or operators that may be applied to move from one state to the next until the goal is achieved.
- Protocol analysis is a key method in problem-solving research, yielding valuable information on the cognitive processes supporting task performance.
- Verbal analysis is another valuable means of analysing verbal data, providing insights into the knowledge representations underpinning task performance.

2 'SIMPLE' PROBLEM SOLVING

In this section we discuss themes and issues in research on what might loosely be termed 'simple' problem solving, although as you shall see, the problems used are not always simple to solve. So-called 'simple' problems, which do not require extensive background knowledge, are sometimes known as 'puzzles' and have often been used in research, as most participants can attempt such problems within a reasonably short time. The issue of representation, and the various ways in which manipulations of problems affect representation, and in turn problem-solving performance, is very much at the centre of this branch of problem-solving research.

2.1 The Gestalt legacy

Simple problem solving began to be studied intensively from the 1910s by a group of German psychologists known as the **Gestaltists**. The hallmarks of the Gestalt approach were the phenomenon of insight, and the view that the whole is greater than the sum of its parts. Insight has famously been labelled the 'aha!' phenomenon, in that sudden restructuring or re-representing of a problem can sometimes lead to a solution.

The Gestalt school particularly emphasized the role of **insight** in problem solving. An example can be found in the story of young Gauss (Hall, 1970) who later went on to become a prominent mathematician (well known for deriving the formula for the normal distribution curve). As a young schoolchild, Gauss surprised his teacher by very quickly producing the correct answer to the sum of all the numbers from 1 to 100. He gave the answer (5050) not by very fast mental arithmetic but by noticing a pattern in the number sequence, viz. that the numbers form pairs ($1 + 100 = 101$, $2 + 99 = 101$, $3 + 98 = 101 \ldots$ and so on). There are 50 pairs and each pair sums to 101, hence the answer is 5050. In this example a good structuring, or representation, of the problem helps considerably.

The processes of restructuring were investigated further by Duncker (1945) who asked participants to think aloud as they tackled problems that required insight to solve. An example is the x-ray problem (see Figure 10.1 below). Participants were shown a diagram and told that it represented a patient with a tumour in the centre of his body. The problem was how to use an x-ray apparatus to destroy the tumour without destroying the surrounding healthy tissue. Participants usually tried alternative restructurings of the problem in terms of sub-goals that could lead to solutions. Thus, the major goal could be achieved if a sub-goal of avoiding damage to healthy tissue could be achieved. The most common solutions involved a sub-goal of lowering intensity of rays on their way through the

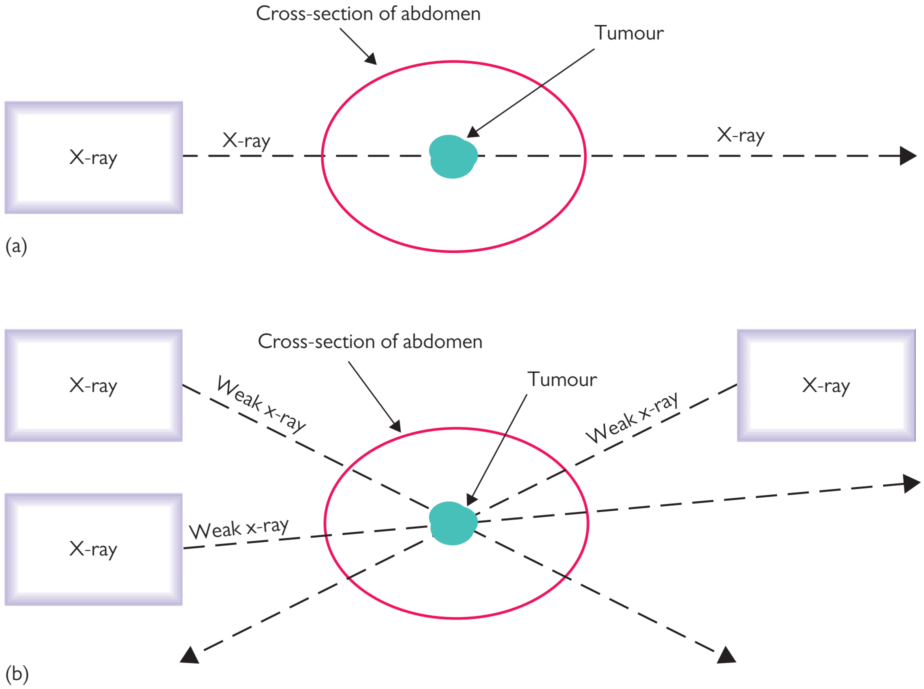


FIGURE 10.1 The x-ray problem.

healthy tissue. This sub-goal led to the solution of using a number of weaker rays, which then converged on the tumour at lethal intensity, thereby destroying the tumour. (An alternative solution involved using a lens to focus a broad band of weak rays on the tumour so that lethal intensity was reached only at the focal point.)

The Gestalt psychologists also investigated cases where insight was generally not achieved because participants were trapped by misleading representations that prevented solution. So-called **'set'** effects arise when learned or habitual ways of tackling a problem prevent the solver from identifying better and simpler methods, or when unwarranted assumptions are made.

Set can be induced by experience with a series of similar problems. Luchins and Luchins (1959) studied problem sets in a series of experiments using water jar problems (presented as a pencil-and-paper exercise). In these tasks participants were asked to say how one could get exactly a specified amount of water using jars of fixed capacity and an unlimited source of water. For example:

> *Given three jars (A, B, and C) of capacities 18, 43, and 10 units respectively, how could you obtain exactly 5 units of water?*

The solution may be expressed as B – A – 2C. After a series of problems with that same general solution, participants had great difficulty with the following problem:

> *Given three jars (A, B, and C) of capacities 28, 76, and 3 units respectively, how could you obtain exactly 25 units of water?*

In fact, the solution to this problem is quite simple (i.e. A – C) but when this problem was presented after a series of problems involving the long solution (B – A – 2C) many participants used the inefficient method and either failed to solve the problem or took considerably longer to use the A – C method than did a control group of participants.

Figure 10.2 illustrates the 9-dot problem, often used to investigate this particular type of set effect. This problem is another example of the set effect, this time produced by the *layout* of the task. Try Activity 10.2.

✱ ACTIVITY 10.2

Starting from any point, draw four straight lines (without lifting the pen from the page) so that each of the nine dots has at least one line running through it.

COMMENT

Most participants interpret the instructions as meaning that they must stay within the square shape of the dots; however, a solution is not possible without breaking this set and going outside the square. Figure 10.3 at the end of the section shows a solution.

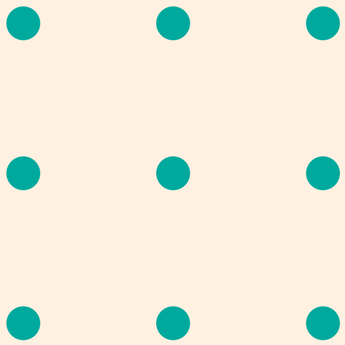


FIGURE 10.2 The 9-dot problem.

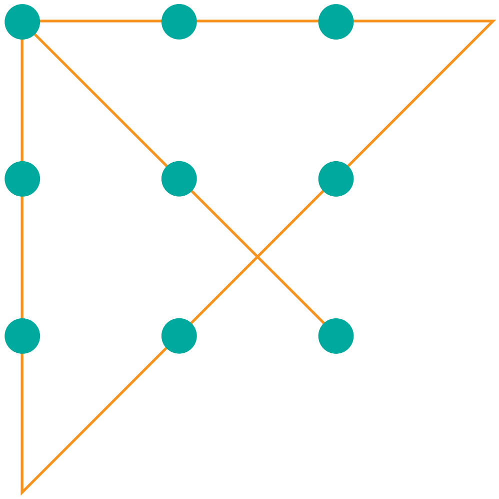


FIGURE 10.3 A solution to the 9-dot problem.

A related block to effective problem solving, known as **'functional fixity'** (also identified by work in the Gestalt tradition), tends to be observed when an object has to be used in a *new way*. Duncker (1945) carried out the classic study of functional fixity using the 'box' (or 'candle' problem). In this task, participants were presented with tacks, matches, three small boxes, and three candles. The problem was to mount the candles side by side on a door, so that they could burn safely. For one group of participants the boxes were empty, but for the other group (experimental group) the boxes were used as containers and held matches, tacks, and candles. The solution is to use the boxes not as containers but as platforms and fix them to the door using the tacks. It was found that the solving rate was much higher in the control group than in the experimental group. Duncker explained this result in terms of a failure to perceive the possible platform function of the boxes when they were presented as containers. Functional fixity has been independently demonstrated and explored further in a number of later studies (e.g. Adamson and Taylor, 1954; Glucksberg and Danks, 1968).

Adults often find it difficult to think of novel uses for familiar objects, as shown in the candle problem. However, young children (5 years old) suffered less from functional fixity in tasks requiring unusual uses of familiar objects, presumably because they had less strongly fixed associations between possible uses and objects (Defeyter and German, 2003). Similarly, Amazonian Indians, unfamiliar with the tools to be used in a task, showed less functional fixity than Westerners for tasks requiring the tools to be used in unusual ways – *until* the typical use was shown to them (German and Barrett, 2005), whereupon they also showed functional fixity at similar levels to Western participants.

Overall, the phenomenon of functional fixity appears to be a robust one and is a likely source of difficulty in real-life problem solving. There are many real-world examples of functional fixity effects. An interesting early example is the history of the steam engine. In 1775, the first Watt steam engine was used to pump water from a colliery, thus solving the problem of flooding. Before steam engines, either buckets or inefficient suction pumps had been used. It was some years before it was appreciated that steam engines could be used for locomotion as well as pumping water.

These early studies demonstrate the importance of representation and its impact upon problem solving. Later research, as we shall see in the next section, went on to examine representational effects in a wider range of problems in more depth.

2.2 Representation in puzzle problem solving

The 'representational effect' has been acknowledged for some time in problem-solving research. Simon and Hayes (1976) constructed several versions of the Tower of Hanoi problem (see Figure 10.4 below), which involves discs of varying sizes arranged on three pegs. The goal is to move the three discs from one peg (e.g. peg A) to another peg (e.g. peg C) using a sequence of legal moves. Typical constraints for this problem are that a larger disc can never be placed on top of a smaller disc (see below, though, for a variation on this rule) and only one disc may be moved at a time.

'State–space' diagrams present a given problem's state at each move juncture. Problem structures can be mapped out and compared using state–space diagrams. (The state–space diagram for the Tower of Hanoi is quite complex. See Figure 10.5, though, for another example of a state–space diagram, drawn to illustrate the structure of the simpler 'hobbits and orcs' problem, previously known as the 'missionaries and cannibals' problem).

Problems that share the same underlying structure (i.e. have identical state–space diagrams) are said to be **isomorphic**. Simon and Hayes contrasted two structurally similar versions of a 'monster' problem. Their 'monster' problem was itself isomorphic to the Tower of Hanoi problem. In the 'move' version, differently sized monsters transferred globes of different sizes to each other according to a set of rules. In the 'change' version, monsters differing in size each held a globe, which had to be changed in size to conform to particular rules. Despite being isomorphs, the 'change' problem was considerably harder to solve than the 'move' problem. People seemed to construct rather different representations of the two problems; the representation constructed for the 'move' problem entailed simpler processing operations than that constructed for the 'change' problem. Our ability to perceive similarities between problems, then, may be influenced by the manner in which problems are represented.

Zhang and Norman (1994) developed a theory to account for representational effects with these sorts of problems. They designed a number of isomorphic versions of the basic Tower of Hanoi problem, and explored ways in which different rules influenced problem difficulty. Their theoretical framework distinguishes between **'internal'** and **'external'** representations. Internal problem representations entail a processing and representational burden, because the

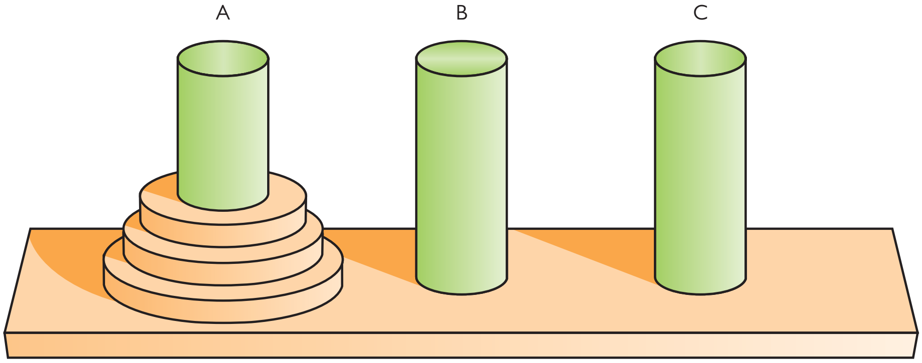


FIGURE 10.4 The three-disc version of the Tower of Hanoi problem.

information needed to solve the problem has to be encoded and maintained in some form. Internal rules, then, are rules that need to be memorized, such as:

1. Only one disc may be transferred at a time.
2. A smaller disc may never be placed on top of a larger disc (notice that this is the reverse of the usual rule for this problem).

External rules differ, however, in that they are not stated explicitly in the instructions but are implied or necessitated by the problem itself. For instance, a form of the Tower of Hanoi where discs are replaced by cups of different sizes filled with coffee involves an external version of Rule 2 above (a smaller cup would fall into the larger cup, spilling coffee). The environment, then, can provide constraints, so some rules need not be internalized. Size and location are properties that need not be internally represented, since differences in size and location are readily perceived. However, dimensional information may be represented internally. For instance, if colour is used to represent some task-relevant information, then the relationships between colours and information may have to be learned.

External representations appear to make problem solving easier, although they also change the nature of the task. We return to this point later in Section 5 in our examination of the relationship between problem solving and learning.

Some problem representations have attributes that may hamper (or facilitate) problem solving. Once a problem has been encoded and represented, problem solving may be described as a search through a set of possible moves (or 'problem space').

2.3 The information processing approach: problem solving as search

Solving a problem may require us to find a suitable sequence of actions drawn from a small set of actions (for instance, moving a series of coloured tiles around a small board until they form a particular pattern or picture). Alternatively, solving some problems may entail selecting or discovering a single action from a large set; for example, using one object that can meet the goal from all objects known to the solver. Within the information processing approach, problem solving is generally seen as a search process.

The initiator of problem solving is a current goal, that is, a representation of a state that is desired but not currently true. Therefore, goals direct the course of thinking by guiding retrieval of goal-relevant material and aiding in the assessment of directions of search as promising or not. Search may proceed in a forwards direction from the starting state by generating possible actions, evaluating the results of those actions, and then choosing those with the best outcomes when assessed against the goal for further exploration. Search may also proceed backwards from the goal by using a problem-reduction or means–ends approach, which breaks down the overall goal into sub-goals that should be easier to achieve.

For example, the problem of booking a trip to New York from London might be broken down (or 'reduced') into sub-goals that could include 'buying a ticket' and 'getting our passports updated'. 'Buying a ticket' can be broken down further into sub-goals such as 'finding an airline' and 'deciding on travel dates' (which in turn can be broken down further still to 'check availability of seats'). The initial goal of getting to New York cannot be achieved until all the steps, and their required conditions, have been identified. The problem is solved by working 'backwards' from the goal, starting with the first sub-goal for which the conditions may be met (e.g. establishing that there are seats available for the date on which we want to travel and selecting these), and then completing other sub-goals until the major goal is achieved and the trip is booked.

A number of studies of search in problem solving have used the Tower of London task (which is similar to the Tower of Hanoi). The problem has a number of variants but basically requires participants to plan out how to move a set of coloured, same-sized discs arranged over three pegs from a starting pattern to a target pattern by moving only one disc at a time. Search processes are assumed to involve the holding of goals and intermediate results in the limited-capacity working memory (see Chapter 9).

Gilhooly *et al.* (1999) studied individuals thinking aloud while solving the Tower of London problem. Their results suggest that working memory limitations tend to shape search patterns so that typically one action is selected from those available at each step. Search builds up a limited length of sequence before returning to the start state and re-exploring. (This process of searching depth first to a certain limit, then backing up and systematically searching all branches of the search tree to the depth limit, is known as 'progressive deepening'.) Gilhooly *et al.* found that the general strategy used was means–ends analysis, which generated a search pattern focused on reducing differences between the current state and the goal state. Means–ends analysis has been generally found to be the typical approach in the related Tower of Hanoi problem (Luger, 1976).

Similar results, indicating very focused mental search, have also arisen from studies of the hobbits and orcs task (Thomas, 1974; Simon and Reed, 1976) and the water jars task (Atwood and Polson, 1976). The state–space diagram for the hobbits and orcs task (Figure 10.5) appears later in Box 10.2.

Means–ends analysis typically involves reducing differences between the current state and the goal state, and so moves that bring the solver closer to the goal tend to be preferred. Thomas (1974) found that participants solving the hobbits and orcs problem found the transition from State 110 to State 221 (see Figure 10.5) especially problematic. The move involves bringing back one hobbit and one orc, which seems at odds with the general strategy of moving closer and closer towards the goal.

The water jars problem has also been used to examine search in problem solving. The problem requires participants to find a way of moving water between jars of given capacities, from a starting state in which the largest is full, to a goal state in which the water is distributed in a particular way over the three jars.

Response data from both the hobbits and orcs and the water jugs problems suggest a model in which solvers look ahead, evaluating a few possible steps at each point in terms of whether they appear to lead to new states closer to the goal or not. For example, at State 331 in Figure 10.5, a solver could look ahead to the next move and evaluate two possibilities, which would move them either to State 220 or State 310. Solvers appear to prefer moves that seem to take them closer to the goal, repeating the entire procedure until the goal is reached (Jeffries *et al.*, 1977). The preference for new states is a heuristic to avoid looping (or revisiting old states in a particular sequence or cycle), as it is all too easy in these tasks to go round in circles! Avoiding loops is also important in the Tower of London task, and Davies (2000) found that participants did not simply rely on memory to judge whether a state had been encountered previously, but would seek to infer whether it could have been a precursor of the present state. If the state in question could not have been a precursor then it was safe to assume that it was new.

2.4 Information processing approaches to insight

Sometimes, a change of representation may be required in order to solve problems that initially induce unhelpful ways of representing the problem. As we discussed earlier, the Gestalt school in the 1920s and 1930s regarded such restructuring as the basis of insight into difficult problems. Though the Gestalt approach was criticized for a lack of clarity in explaining how restructuring took place, Ohlsson (1992) offered some suggestions as to how restructuring might occur.

Ohlsson (1992) proposed that when working on a problem people generate possible actions or operators from long-term memory, which are cued by the problem representation. Applying the operators to the current problem state leads to a new problem state, which in turn elicits further possible operators. In this way a problem with a useful initial

BOX 10.2 The hobbits and orcs task

The hobbits and orcs task requires participants to find a way of transporting three hobbits and three orcs safely across a river in a boat. The boat can only hold two creatures at a time and on either side of the river the orcs must never outnumber the hobbits at any time.

In this state–space diagram (see Figure 10.5), each box represents a single state of the task. The number of hobbits (H) and orcs (O) on the left- and right-hand side of each box indicates the number of hobbits and orcs on the left and right banks of the river at any given time. Each state is labelled with a three-digit number, with the first digit representing the number of hobbits, the second the number of orcs, and the third the number of boats all on the left bank of the river. So, for example, State 331 (near top left) indicates the start of the problem, with three hobbits, three orcs and the boat all on the left bank of the river. State 000 (top right) is the solution state, with all six creatures and the boat transported to the opposite side.

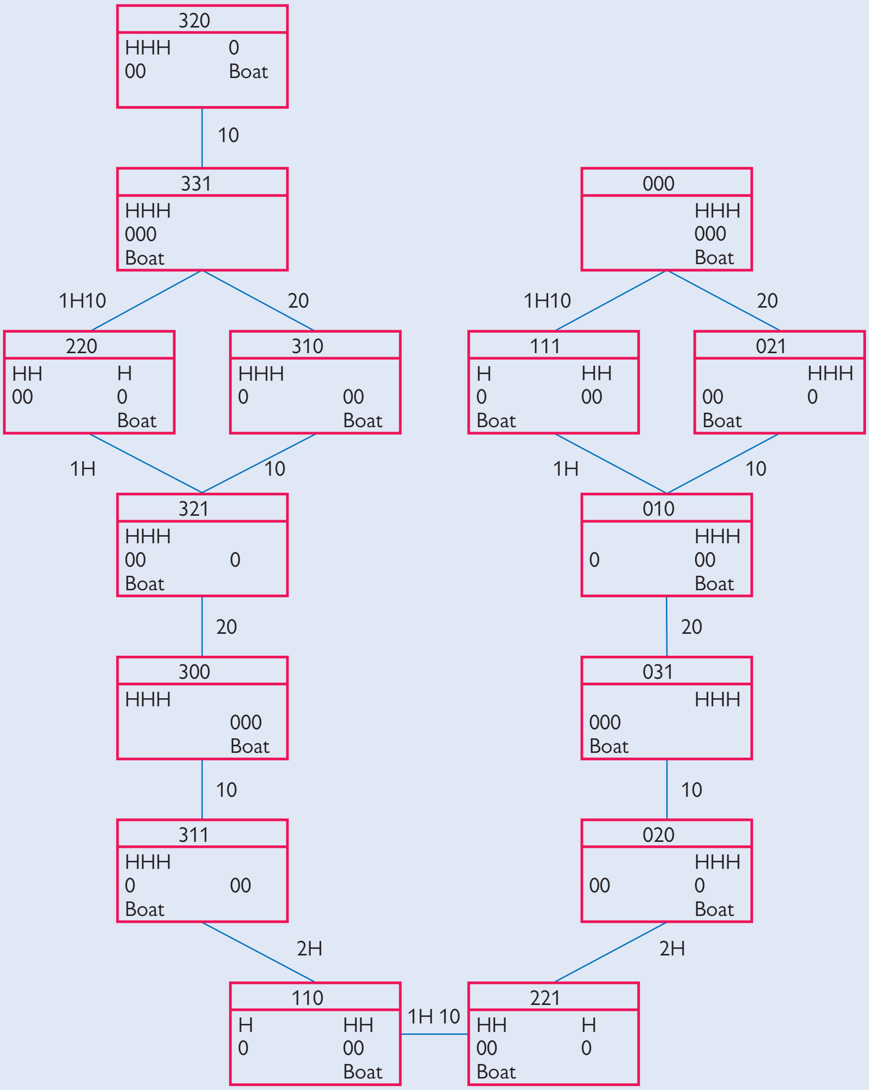

320
HHH 0
00 Boat
10
331
HHH
000
Boat
1H10
20
220
HH H
00 0
Boat
310
HHH
0 00
Boat
1H
10
321
HHH
00 0
Boat
20
300
HHH
000
Boat
10
311
HHH
0 00
Boat
2H
110
H HH
0 00
Boat
1H 10
221
HH H
00 0
Boat
2H
020
HHH
00 0
Boat
10
031
HHH
000
Boat
20
010
HHH
0 00
Boat
1H
10
111
H HH
0 00
Boat
021
HHH
00 0
Boat
1H10
20
000
HHH
000
Boat

FIGURE 10.5 State–space diagram for the hobbits and orcs problem.

representation may be solved, as eventually a state is reached from which cued operators lead to the goal. However, if the initial representation is misleading, then a state will be reached in which no new useful operators can be retrieved. Subjectively, we experience this state, labelled an **impasse,** as a mental 'blank'; we cannot think of anything new to try. Ohlsson suggests that impasses can be overcome by changing the problem representation so that new operators can be cued, retrieved, and tried out. Specifically, Ohlsson proposed three ways in which the problem representation could be changed or restructured: (1) **elaboration**; (2) **re-encoding**; and (3) **constraint relaxation**. Together, these form the basis for 'Representational Change' theory, which we examine more closely next.

Elaboration involves adding information to the representation by observing previously unnoticed features. For example, to use the matchbox tray in solving the 'candle' problem (Duncker, 1945) discussed earlier, the solver has to notice the possible use of the tray as a platform.

Re-encoding involves changing the encoding rather than simply adding new information. For example, how could a man have legally married 20 women in one month in a country where polygamy is illegal and none of the women have died or been divorced? To solve this, we have to change the encoding of 'married' from the man becoming husband to each woman to the man causing others to become married to each other; for example, the man in question could be a minister who is entitled to perform marriage ceremonies.

Constraint relaxation involves making the goal requirements less restrictive than initially assumed. For example, one source of difficulty in the 9-dot problem is the tendency to over-restrict the goal so that the four lines are kept within the square array of dots; removing (or relaxing) this constraint is necessary for a solution. We return to the 9-dot problem in Section 2.5.

A further demonstration of the role of constraint relaxation comes from studies by Knoblich *et al.* (1999) of matchstick algebra problems. Examples of these types of problem are as follows: make the equations below, involving Roman numerals, true by moving a single stick (see Figure 10.6).

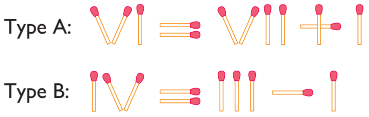


FIGURE 10.6 Matchstick algebra problem.

There are large differences in solution times and solution rates for these two types of problem, with Type A being easier than Type B. In the Type A example, the rightmost stick from the 'VII' is moved to the right of 'VI'. This gives 'VII = VI + I', which is true. In the Type B problem, the operators '=' and '–' are changed by moving one of the horizontal sticks from the '=' and placing it over the '–' to make the true equation 'IV–III = I'. Knoblich *et al.* argue that it is harder to break the constraint on changing operators than on changing number values.

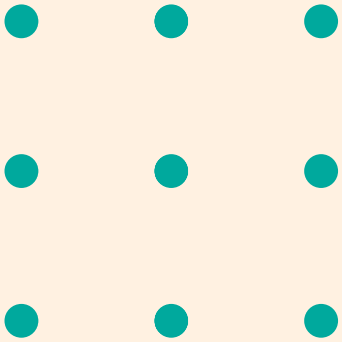
2.5 Insight: inappropriate heuristics or misleading representations?

MacGregor *et al.* (2001) have developed an alternative to Representational Change theory known as 'Progress Monitoring theory' (or sometimes as 'the Criterion for Satisfactory Progress theory'). According to this approach the main source of difficulty in insight tasks is the use of inappropriate heuristics (particularly **hill-climbing** heuristics, in which solvers always choose the action that leads to the most immediate progress towards the goal). They propose that failure to meet a progress criterion triggers restructuring rather than impasses. MacGregor *et al.* (2001) applied their approach to the 9-dot problem, which to remind the reader is shown again in Activity 10.3.

Traditional explanations for the difficulty of this task propose a fixation (set) on the square shape, so that other solutions are not considered. However, instructions to search outside the square were not found to be helpful (Weisberg and Alba, 1981). Lung and Dominowski (1985) suggested

ACTIVITY 10.3

Connect the nine dots with four connected straight lines without lifting your pencil from the page as you draw.



other inappropriate constraints, e.g. assuming all lines begin and end with dots.

Progress Monitoring theory proposed an alternative explanation involving two main points. These are: (1) use of a maximization heuristic, in which each move or decision is an attempt to make as much headway as possible towards the goal; and (2) use of progress monitoring, in which the rate of progress is assessed constantly, and if it is deemed to be too slow and inefficient then criterion failure occurs. An alternative strategy may then be sought.

Progress Monitoring theory as applied to the 9-dot task suggests: (1) that the maximization heuristic would be for each move to cover as many new dots as possible; and (2) that progress monitoring would involve the rate of progress being assessed against the number of dots required to be covered per line to solve, and if no move meets the criterion, criterion failure occurs. An alternative strategy may then be sought, e.g. extending the lines beyond the square.

MacGregor *et al.* explored the Progress Monitoring theory explanation of the 9-dot task by testing participants with two variants of the problem labelled version A and version B, as shown in Figure 10.7.

If constraint relaxation is all that is required to think ‘outside the box’, then participants should do better on version A than on version B, since A shows a line going out of the box. However, if criterion failure is necessary then participants will do better on B, because they can cover fewer dots in the next two moves than are needed for solution, and so will realize that they are on the wrong path sooner. MacGregor *et al.* found that only 31 per cent of those given version A were successful, while 53 per cent of those given version B solved it successfully.

Further experiments on Progress Monitoring theory used coin manipulation problems such as the 8-coin problems shown in Figure 10.8, in which people have to move only two coins so that each coin is left touching exactly three others. Figure 10.8 depicts two different arrays of coins, and for each array the task is the same.

If the strategy employed simply seeks to achieve a short-term goal of bringing one particular coin to rest

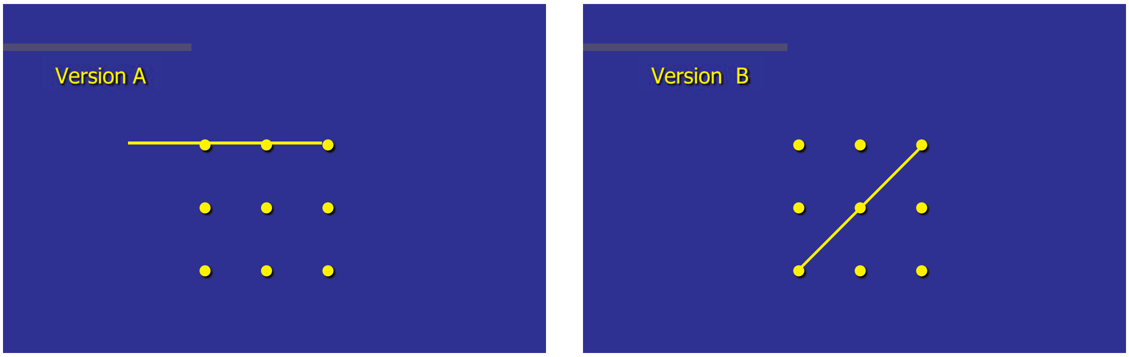


FIGURE 10.7 Progress Monitoring theory: MacGregor *et al.*’s version A and version B of the 9-dot problem.

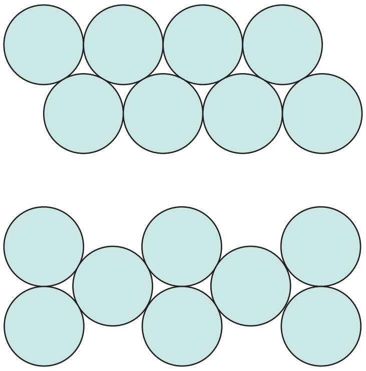


FIGURE 10.8 Ormerod *et al.*'s (2002) 8-coin problem. Moving only two coins, leave each coin touching three others.

in contact with three others, then there is no two-dimensional move available in the upper version of the problem, but 20 two-dimensional moves available in the lower version that put the moved coin in touch with three others. Thus, criterion failure will be reached much sooner in the upper version and so more solutions should result. In the lower version a lot of effort would be wasted in exploring what look to be promising moves, but would not lead to a solution. As predicted by the theory, 92 per cent solved the problem in the upper version compared with only 67 per cent in the lower version. The solution is shown in Figure 10.9.

Overall, the central claim of Progress Monitoring theory is that insight is most likely to occur when constraint relaxation follows criterion failure. There is good evidence for this from the above studies. So, the theory deals well with the motivation for changing strategy but it is less clear about how new strategies are actually reached.

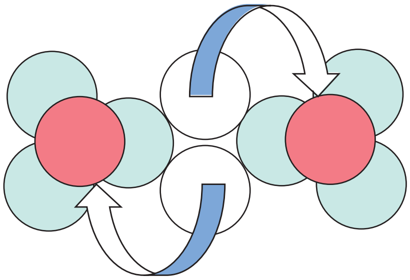


FIGURE 10.9 Solution to the 8-coin problem: moving only two coins leave each coin touching three others.

2.6 Testing two insight theories

Jones (2003) compared Representational Change and Progress Monitoring theories using car parking problems, in which toy cars in a car park must be moved around to let a toy taxi out of the car park. The cars can only be moved back and forwards in straight lines and cannot be picked up. The problems varied so that some were simple and would not induce impasses, restructuring, or insight, while others were more difficult and did tend to produce impasses and, subsequently, insight. The main difficulty in the insight inducing problem is that participants need to consider moving the taxi before the exit path has been cleared. Two example problems are shown in Figure 10.10; the first is easier than the second, which tends to produce impasses and insight.

Jones's (2003) experiment involved three conditions. In the Normal condition, four progressively harder (more moves) problems were followed by an insight task. The Rotated condition was the same, but the insight task was rotated by 90 degrees, which would reduce set effects on the Representational Change theory. In the Easy condition, four easy problems were followed by an insight problem. A number of dependent variables were examined, viz. moves, times per move, and location and duration of eye fixations. The Representational Change theory predicted that impasses would occur before the taxi was moved and that the Rotated and Easy conditions would lead to better performance on the insight problem than the Normal condition. Progressive Monitoring theory predicted that there would be no difference between the Rotated and Easy conditions and that early impasses would be associated with better performance (because greater look-ahead would lead to earlier impasses). Results indicated that impasses (indicated by long fixation times) did

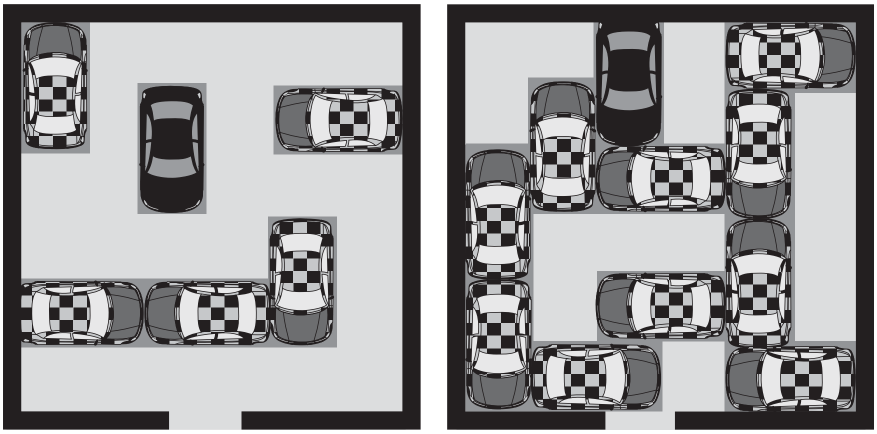


FIGURE 10.10 a) Taxi problem: move the cars to let the taxi out of the parking area. b) More complex taxi problem: move the cars to let the taxi out of the parking area.

tend to precede moving the taxi, and that Rotated and Easy conditions did lead to better performance than the Normal condition. Further, earlier impasses were associated with better performance on the insight task.

In conclusion, both Representational Change and Progress Monitoring received some support from Jones (2003). The results suggested that Progress Monitoring applies to the first stage of solving, leading to impasse, and that Representational Change applies to breaking of impasse and subsequent insight. Thus, it seems that both theories have support and that the two theories deal with different stages of insight solving, so are complementary rather than contradictory.

SUMMARY OF SECTION 2

- Problem solving begins with a problem representation.
- The information processing approach analyses problem solving in terms of search within the space of possibilities arising from a particular way of representing the problem.
- Manipulating instructions and the appearance of problems appears to influence the nature of the problem representation constructed, which in turn can affect the ease with which the problem may be solved.
- Some representations of supposedly 'simple' problems render the problems extremely difficult to solve.

3 ANALOGICAL PROBLEM SOLVING

As we saw in the previous section, some problems we encounter are novel and difficult to solve even when they require a minimum of background knowledge and experience. However, we often encounter problems that are rather similar to problems we have tackled previously. Even if a solution to a new problem is not known, we may know and be reminded of the solution to similar problems and be able to use that known solution to suggest a solution to the new problem. That is, we may be guided to solution by the use of an analogy. Whilst other animals make use of association and similarity, it is arguably the facility with which humans perceive and use relational similarity that sets us apart. In this section, we examine some of the reasons why analogical thinking is so important.

3.1 Analogies in problem solving

Spellman and Holyoak (1992) found that experimental participants readily accepted analogies of the kind often used in discussions of international politics. For example, when Saddam Hussein attacked Kuwait in 1990, many commentators likened him to Adolf Hitler and George Bush Sr. to Winston Churchill. Similarly, some regarded the 'Domino' theory – the US government's belief that if one Asian country fell to communism then others would quickly follow – as compelling justification for the war in Vietnam.

In science, analogies have often been used to develop understanding. For example, the heart has been seen as analogous to a water pump and atomic structure as similar to that of the Solar System. In cognitive psychology, we hope that analogies between human and computer information processing will likewise prove useful.

Studies of analogy use in problem solving have often used Duncker's x-ray problem, which you met in Section 2, as the target problem. You may remember that the problem involved using rays to destroy a tumour in the centre of a body without destroying healthy tissue. The solution was to have a number of weak rays converge on the tumour, which would then have a cumulative effect at that point. Gick and Holyoak (1980) gave their participants an analogous (or 'base') story about a general seeking to take a castle who had to divide his forces into small groups, which then attacked the castle simultaneously on all sides. Participants were then given the x-ray problem, with and without a hint. The hint indicated that solvers should try to use the story to solve the x-ray problem. A control group did not receive the analogue before tackling the x-ray problem. Rate of solving was low for participants who had not been given the analogue, somewhat higher when the analogue alone had been given, and markedly higher for the analogue followed by a hint.

Later studies (Holyoak and Koh, 1987; Keane, 1988) indicated that the closer the base story is in surface features to the target problem, the more transfer is likely. For example, Keane found that a very close analogy of a surgeon treating a brain tumour by radiation was much more often retrieved and used in tackling the x-ray problem, even after a week's delay, than the more remotely analogous story of the general dividing his forces.

Anolli *et al.* (2001) found that retrieval of a remote analogy was ineffective in itself without provision of a hint that the analogy contained a useful clue to the solution. In a series of seven studies, little benefit in terms of solving was found for simply reminding participants of the analogous story without a hint to use the story. However, retrieval plus a hint was very effective.

Dunbar (2001) has noted that there is an *analogical paradox*. The paradox is that in real life, use of abstract analogies that depend on deep structural similarities is common, while in laboratory studies, participants tend to use superficial features and have difficulty with deeper forms of analogy. Blanchette and Dunbar (2000) found that participants who were asked to produce analogies that could be used in arguments about whether or not drastic cuts should be made in public spending during a budgetary crisis readily produced deep analogies that drew on a range of content areas having little superficial resemblance to political matters. Blanchette and Dunbar proposed that generating analogies requires participants to use structural rather than superficial features; also, the subject matter of the naturalistic analogy studies has been familiar and understood in some depth. In typical laboratory studies, on the other hand, the material is not highly familiar and there is little pressure to encode the base story in a deep way. A possible interpretation of Anolli *et al.*'s (2001) results (namely, target problem solution is increased by reminding of the base story plus a hint that the analogy contained a useful clue to the solution) is that the hint encouraged a deeper structural representation of the base story, which in turn facilitated application of the analogy.

3.2 How do analogies work?

The first detailed theory of how people apply analogies in problem solving was the 'structure-mapping' theory (Gentner, 1983; Gentner and Markman, 1997; Gentner *et al.*, 2001). According to this theory, there is a process of **analogical mapping** whereby a **structural alignment** is established between the representations of the base and the target. That is to say, explicit correspondences are established between the represented elements and relationships in the two situations. As an example, consider the Solar System analogy of the atom. Typically, the Solar System would be represented as having two types of object (the Sun and the planets) and these exhibit various properties and relationships (e.g. the Sun is more massive than the planets; the planets orbit the Sun). The analogy would align the Sun with the nucleus of the atom and the planets with electrons. Aspects of the Solar System model that do not map are omitted (e.g. no equivalent of moons in the atom). Higher-order relationships also guide the alignment process. Thus, there is a higher-order relationship such that less massive objects orbit more massive objects; this guides the inference from the Solar System analogy that the electrons revolve around the nucleus. Falkenhaimer *et al.* (1986) have successfully implemented the detailed model as a computer model known as the 'structure mapping engine' (SME).

Gentner and Gentner (1983) showed that different base analogies were common for understanding electrical flow. These influenced how well people solved different problems regarding electrical flow through circuits with batteries and resistors arranged in parallel or in series. The main analogies were electricity as a flow of fluid through pipes or as a flow of crowds through passageways. Example differences between the two analogical mappings are that in the fluid analogy electrical resistance would be mapped to pipe width, while in the crowd analogy resistance is mapped to gates in the passageway. People who used the fluid analogy performed better on battery problems; people who used crowd analogies were better on resistor problems. People pre-trained on one or other of the analogies also showed similar results.

A curious and almost counterintuitive aspect of the process of analogical mapping is Gentner's (1983) claim that this is symmetric at the outset, though the inferences that are subsequently projected may be directional. Thus, Gentner's position would argue that in the initial stages of processing, 'my legs are like lead' would be as comprehensible as 'lead is like my legs'. An alternative position, associated with Glucksberg *et al.* (1982) for example, asserts that metaphor processing is directional from the beginning. If this were the case, then reverse metaphors such as 'lead is like my legs' should not be comprehensible. The debate hinges on whether metaphor use is based on comparison processing or on category processing (see Chapter 5). A metaphor may be described in terms of the 'vehicle' and the 'topic'. Take for example the metaphor *my child is a monster*. In this case, the topic is 'my child' and the vehicle is 'a monster'. Gentner argues that in understanding this metaphor, we compare the topic and base representations and then arrive at a common relational system: 'My child is bad, just as monsters are bad'. For Glucksberg *et al.*, in using a metaphor one asserts that the topic ('child') is a member of the category of which the vehicle ('monster') is a prototypical member. An issue for category theories is what incites the listener to create a metaphorical category for the vehicle, rather than using its literal meaning.

One way to test the symmetry hypothesis is to use a variant on Glucksberg *et al.*'s (1982) metaphoric interference technique, which requires participants to state whether metaphors are comprehensible or not. Wolff and Gentner (2000) used four types of statement: false statements (e.g. some cats are peanuts); forward metaphors (e.g. some jobs are prisons, some builders are cowboys); scrambled metaphors (e.g. some jobs are cowboys); and reverse metaphors (e.g. some prisons are jobs). Participants verify the statements as either true or false after either a 600 ms or a 1200 ms exposure. Typically what is found is that the standard, forward metaphors interfere with true-false judgements, these judgements being slower than judgements of the literal false statements. One interpretation of this is that metaphor processing occurs very early on. If metaphor use is based on categorical processing as Glucksberg *et al.*'s finding suggests, then we should see the influence of both categorical information associated with the vehicle, and attributes of the topic, but only for the standard, forward metaphors. Reverse metaphors would simply be anomalous and verification times should be comparable to false statements and to scrambled metaphors. If, on the other hand, metaphor processing is symmetrical at the outset, then reverse metaphors should show similar levels of interference to forward metaphors. Results supported the symmetry position by showing that the interference effect was comparable for forward and reverse metaphors, providing further support for the structure-mapping theory.[2] Interestingly too, the exposure manipulation corroborates this conclusion. Individuals were equally as fast in comprehending forward and reverse metaphors after 600 ms, and it was only after 1200 ms of exposure that forward metaphors were judged more comprehensible than reverse metaphors.

Underpinning our ability to engage in analogical thinking is our capacity for relational reasoning. Relational concepts must be learned, and our capacity to acquire and use relational concepts is augmented by language. Gentner (2003) notes that enhanced relational encoding may be brought about by the acquisition of a more uniform relational structure (reflected in vocabulary use, for example), and this in turn supports relational retrieval. Rattermann and Gentner (1998) showed that teaching 3-year-olds how to use relational labels (such as big, tiny, little) enabled them to reason as well as 5-year-olds who already understood and used such labels.

SUMMARY OF SECTION 3

- Research on analogical problem solving illustrates a paradox: we often cannot help but be reminded of problems similar to one we presently face. However, we often fail to see the crucial relationships between a current problem and one we have previously encountered.
- Representation seems to be at the core of the paradox.
- The mapping of correspondences between target and base constituents appears to be a very early stage in the analogical mapping process, which is symmetrical.

4 'COMPLEX' PROBLEM SOLVING

While research on the ways in which we solve puzzles and analogies has mapped out the terrain to a certain extent, it should be apparent that understanding how people solve the relatively 'knowledge-lean' problems we have looked at so far is only a part of the picture. Many problems require a considerable amount of knowledge if they are to be solved successfully. This section focuses on 'complex' problem solving, or problem solving that requires an extensive knowledge base. Does knowledge of a domain affect problem representation, and hence the likelihood that a problem will be solved? Do problem representations change as knowledge is acquired and as skill develops? Can we characterize the development of skill in problem solving? Researchers have turned their attention to how experts and novices solve problems in their attempts to try to answer these questions.

While it may seem obvious that experts know more than novices, until relatively recently the layperson's view of the expert might well have encapsulated the view that experts owe their skill to superior mental capacities, rather than to a vast body of specialist knowledge. The shift in emphasis began with some ground-breaking research on chess skill. We shall examine this work in some detail, because much of the later research on expert and novice problem solving developed from the early chess studies, and because findings obtained in later studies tend to echo those of the chess experiments described below.

---

[2] A number of models similar to the structure-mapping theory have been proposed, including the 'analogical constraint mapping engine' (ACME) (Holyoak and Thagard, 1989) and the 'incremental analogy machine' (IAM) (Keane, 1994).

First, some words on terminology. Researchers have examined both adversary and non-adversary problem solving. Chess play is an example of adversarial problem solving, because the game of chess involves an opponent. Code-breaking, de-bugging computer programs, and medical diagnosis are examples of non-adversarial problem domains. Those engaged in adversary problem solving must consider not only their own possible actions, but also those of an opponent.

4.1 The role of knowledge in expert problem solving

4.1.1 Early chess studies

De Groot (1965; original 1946) carried out a series of now classic studies of chess players. These were extremely significant and heralded the start of a new emphasis on knowledge in skilled problem solving. Information processing had taken centre stage as the dominant paradigm and many researchers busied themselves with the construction of models of cognitive processors and processes. Up until then, it had been implicitly, if not explicitly, assumed that skilled problem solvers must have superior information processing capabilities. De Groot tested this assumption in a novel way.

De Groot asked five grand masters (the highest skill level attainable in chess) and five skilled players to think aloud as they studied a chessboard and chose a move. If information processing capacities are a key determinant of expertise, then we would expect to find the grand master players, with their superior capacities, searching further ahead and conducting broader searches for candidate moves. The evidence from the think-aloud protocols, however, was surprising and revealed no reliable quantitative differences at all between the grand masters and the highly skilled players. The only difference that did emerge between the two groups was unremarkable – the grand masters ultimately chose better moves.

De Groot also employed what is known as a 'recall-reconstruction' paradigm (see Box 10.3). He showed chess players chessboards with pieces arranged from actual games. The boards were presented to players for 2–15 seconds and then removed. He then asked the chess players to reconstruct the board positions from memory. The chess masters could reconstruct the boards almost without error (91 per cent of pieces correctly replaced), whereas the poorer players averaged only 41 per cent correct. Skill level was then linked to the amount of information remembered about the chessboard positions.

Chase and Simon (1973b) devised a second task, where players had to reconstruct a chessboard while the board they had to correctly match was still in view. Although this may seem an odd task, the point was to find out how many pieces were placed on the target board after each glance at the board they were copying, what those pieces were, and how much time elapsed between placing pieces on the board. Chase and Simon found that the strongest chess players replaced more chess pieces on the board following each glance than the weaker players did, replaced pieces more quickly, and were more likely to replace pieces together that bore some meaningful relationship to each other. These findings suggested that experts not only possess more knowledge about their domain of expertise, but that their knowledge is organized in more meaningful and readily accessible ways.

These early studies of chess skill showed that skill depended at least in part on the acquisition of domain knowledge, and stimulated a vast amount of research on the nature of expert problem solving and the relationship between knowledge and skill. We summarize some of the key studies below. These studies sought to characterize the empirical phenomena

BOX 10.3 RESEARCH STUDY The recall-reconstruction paradigm in chess

Chase and Simon (1973a) extended the basic chessboard recall-reconstruction paradigm, originally used by Lemmens and Jongman in an unpublished study in 1964. In one study, Chase and Simon presented boards with between 20 and 22 chess pieces arranged on them to three chess players (a master, a class A (highly skilled) player, and a novice). Some of the boards were presented with chess pieces arranged as they might be in a real game (see Figure 10.11), while others were presented with the chess pieces arranged at random.

Players were given five seconds to study each board. The board was then covered up and each player was asked to reconstruct on another board the position just seen. Results from the first memory trial showed that the master player was much better (16 pieces replaced correctly) at accurately replacing the chess pieces than both the class A player (8 pieces replaced correctly) and the novice player (only 4 pieces replaced correctly). The skilled players' advantage only held for chessboards with pieces placed in plausible, real-game positions. When the different players were asked to reconstruct the random boards from memory, all players performed equally poorly, correctly replacing only a small proportion of pieces. This suggested a connection between memory for meaningful patterns and problem-solving skill.

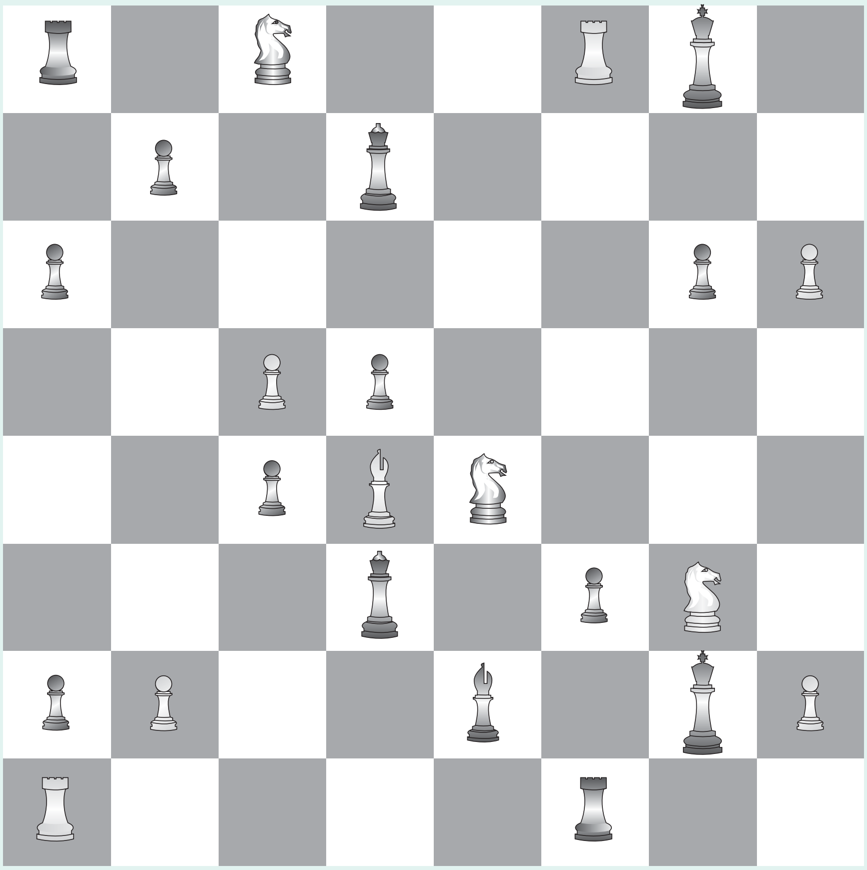


FIGURE 10.11 A chessboard position from a real game. *Source: adapted from Chase and Simon, 1973a*

associated with skill in problem solving, phenomena that theories of skill acquisition and problem solving would ultimately have to accommodate and explain.

4.1.2 Experts work forwards

Larkin *et al.* (1980) were interested in possible strategic differences between experts and novices. Experts know more than novices, but do they also use qualitatively different problem-solving strategies from those used by novices? They asked expert and novice physicists to solve a range of physics problems. Using protocol analysis, they found that experts tended to use a working forwards strategy, beginning with information given in the problem statement and using that to derive a solution. Novices, on the other hand, used a working backwards strategy (means–ends analysis), starting with the goal, or quantity to be solved, and working backwards from that to the given information until they were able to solve one part of the problem. Novices then typically retraced their steps, working forwards until the problem was solved.

Why do experts and novices use different problem-solving strategies? It appears that experts use their domain knowledge to generate a good problem representation, which supports the use of a working forwards strategy. In the absence of detailed knowledge about the relationships between variables relevant to the problem, novices seem to have no option but to fall back on means–ends analysis, or even trial and error learning.

4.1.3 Experts have better problem representations

Chi *et al.* (1982) tackled the issue of problem representation and categorization by experts and novices. Experts know more and they use distinct problem-solving strategies. Is expert problem solving also supported by more effective ways of representing and categorizing problems? Chi *et al.* asked expert and novice physicists to think aloud as they categorized physics problems on the basis of similarities in terms of how the problems might be solved. Unusually, then, participants did not actually have to solve the problems.

The two skill groups did not differ on quantitative measures, such as number of categories or time to categorize. This showed that novices were not limited in their capacity to discriminate problems. However, there were clear qualitative differences in the nature of the categories into which problems were sorted. Novices referred to objects and key words contained in the problem (such as 'levers' and 'pulleys'), and appeared to use these irrelevant 'surface structure' details as a basis for categorization. Experts, on the other hand, referred to the physics principles and laws (the 'deep structure') that were needed to solve the problems in their justifications. Problems that could be solved by reference to the same principle or law were perceived by the experts to be similar and were grouped together. Novices tended to group together problems that were similar in 'surface structure', while experts sorted problems on the basis of similarity in 'deep structure'. It seems then that experts are aware of commonalities between problems in terms of how they might be solved.

Schoenfeld and Herrmann (1982) carried out a rather similar study, looking at mathematical problem categorization among mathematics professors and novices. Their participants read through the set of problems and then grouped together those problems they considered to be mathematically similar. The study confirmed the findings of Chi *et al.* (1982), with novices sorting the problems on the basis of superficial details, or surface structure, and the professors sorting problems on the basis of similarities in solution methods, or deep structure. (You will notice some similarities here with the discussion in Chapter 5 of how different groups of people categorized trees.)

Chi *et al.* found in their study that experts were able to perceive an appropriate solution method within 45 seconds. This suggests that knowledge useful for a particular problem is accessed, or becomes available, when a problem is categorized as a specific type. These categories may correspond to problem schemata or 'packets' of knowledge that can be used to solve a particular type of problem.

4.1.4 Experts become expert through extensive practice

It is often said that 'practice makes perfect'. In the context of problem solving, researchers noticed many years ago that performance improves with practice in a very systematic and predictable fashion. What is particularly interesting is the observation that, regardless of what is being learned, performance improves with practice in a highly predictable way. The relationship, known as the 'power law of practice', has been known for a long time, though there is an ongoing debate as to whether practice learning data are best fit by a power function, or some other function. The relationship shows up in Snoddy's (1926) study of mirror-tracing of visual mazes. It appears in perceptual tasks such as Kolers' (1975) studies on mirror-reading (where text is transformed), in pattern recognition (Neisser *et al.*, 1963), and in tasks from the domain of human–computer interaction (e.g. Card *et al.*, 1983).

Practice therefore seems to be a factor in the development of skill. The improvement in performance with practice applies over a wide range of activities that are better described as 'tasks', and that include problem-solving tasks as well as other kinds of tasks (for example, juggling and search tasks). Why does performance improve with practice? Three main classes of explanation have been proposed:

1. Individual task components are executed more efficiently.
2. Sequences of task components are executed more efficiently.
3. Qualitative changes occur in representations of task structure.

The first two explanations argue that performance improves with practice because the piecemeal recovery of declarative knowledge into working memory is reduced, and because we learn to run off sequences of procedures in ever greater units or chunks. The third explanation asserts that performance improves because the nature of the task changes, either because the task is restructured or because we shift from algorithm-based to memory-based processing (an example of the latter is Logan's (1988) 'instance' theory of automaticity).

How much practice is needed to achieve excellence? Ericsson *et al.* (1993) have given ten years as a ballpark figure for attaining high levels of performance in a variety of areas (e.g. chess, mathematics, and violin playing). In a review of the literature on practice and performance, Ericsson (1991) has suggested that it takes at least ten years to reach the international level of performance in sports, the arts, and sciences. Simon and Chase (1973) estimated that it took some 3,000 hours of practice to become an expert and around 30,000 hours to become a chess master. The preparation period may often commence at an early age, possibly because it takes so long to acquire the necessary knowledge. While it clearly takes a long time to attain very high levels of performance, it is nevertheless possible to train subjects to improve on their previous best performance. Ericsson and Harris (1990) trained an individual who was not a chess player over a period of 50 hours to recognize chess positions almost as accurately as some chess masters.

However, as Ericsson and Polson (1988) found, practice itself is not a guarantee of superior performance. In their study, the waiter most skilled in remembering orders used more effective encoding strategies and achieved much better performance than his equally experienced counterparts, who did not use the same optimal encoding strategies to remember dinner orders. This means that something else must mediate between practice and performance.

What appears to be critically important is not just how much practice individuals have, but what they actually do while they are practising the skill. If it takes a very long time to become expert, then clearly we need to document what individuals do over a longer time scale than is usually considered. We return to this point in Section 5.2 when we explore individual differences in problem-solving performance.

4.2 A modal model of expertise?

The early chess studies triggered a vast amount of research that used what became known as the 'expert–novice' paradigm. The model of chess expertise that emerged became known as the 'pattern recognition hypothesis', because it assumed that skilled performance depended upon the ability to access previously learned patterns, such as configurations of chess pieces on a board, from long-term memory.

The general idea that performance depends upon a large body of highly structured domain-relevant knowledge and skill has been borrowed by researchers examining skill-related differences in non-adversarial domains such as physics (Chi *et al.*, 1982), mathematical problem solving (Schoenfeld and Herrmann, 1982), computer programming (McKeithen *et al.*, 1981), and political science problem solving (Voss *et al.*, 1983). Results consistently showed a link between expertise and knowledge, suggesting that a 'modal model' of expertise was emerging, whereby expertise depends upon the acquisition and organization in long-term memory of domain-relevant knowledge and skill.

Although supported by the data, these initial observations about expertise seem descriptive and lack explanatory power. Sternberg (1995) is just one of the researchers to have commented upon this. Over-use of the paradigm appeared to constrain the nature of the findings to a series of observations about experts 'knowing more' than novices. Accounting for these findings was nonetheless a challenge for theories of skill acquisition, but many researchers recognized that there was more to expertise than the gradual accumulation of domain-specific knowledge.

As we shall see in the next section, when researchers began to explore different questions about the nature of expertise and about skill acquisition, some findings emerged that challenged the prevailing view of expertise while at the same time yielding valuable insights into the relationship between memory and skill, and the development of expertise itself.

SUMMARY OF SECTION 4

- In memory tests for information from their domain of expertise, experts recall more than novices.
- Experts are superior to novices in knowledge rather than in basic capacities.
- Experts use a working forwards strategy, while novices tend to work backwards.
- Experts construct better problem representations than novices.
- Experts become expert through extensive practice.

5 PROSPECTS FOR PROBLEM-SOLVING RESEARCH

In this section, we focus on research that points to some limitations of the general model of expertise outlined above, and we go on to discuss some of the directions that research in problem solving has been taking.

5.1 Limitations of the modal model

It is perhaps ironic that early indications that all was not well with the modal model of expertise came from research on chess skill, which had originally played such a large part in stimulating research on expert and novice problem solving.

5.1.1 Chess skill and memory

The classic chess recall–reconstruction experiments (as discussed in Box 10.3) showed that the master chess players' memory advantage held only for meaningful chess positions, suggesting that memory determines chess skill. However, subsequent studies of expert chess play questioned this conclusion and suggested that memory cannot be the sole determinant of skill (Holding, 1985). The key distinction explored then was that between the respective roles of search and pattern recognition in chess skill.

For instance, Holding and Reynolds (1982) sought to determine whether skill differences could be shown in the absence of differences in memory. They asked players differing in their skill ratings to recall semi-random positions after a brief 8-second presentation. Next, players were asked to select the best continuation moves. Skill level was unrelated to recall of semi-random positions, replicating the findings of de Groot (1965). However, the interesting finding was that the number of best moves chosen after a few minutes of consideration correlated positively with playing strength. Therefore, differences in memory for chess patterns could not account for the finding that better players chose more good moves from random starting states. This suggests that, for highly skilled chess players at least, something other than memory for highly familiar configurations of chess pieces may be implicated in chess skill. However, Schultetus and Charness (1999) repeated the Holding and Reynolds study, but with an important alteration. They asked the players to recall the position once the position had been evaluated and the best next move chosen. Like Holding and Reynolds, they found that skill level did not correlate with recall after the brief exposure, but following the board evaluation and move selection task, the stronger players achieved better recall performance than the less skilled players. These results are consistent with the notion of a long-term working memory (Ericsson and Kintsch, 1995). We turn to this in the next section.

5.1.2 Long-term working memory

Skilled performers are able to access stored information and maintain performance levels even after interruption, and this is one factor that gave rise to Ericsson and Kintsch's (1995) notion of long-term working memory. Unlike working memory (see Chapter 9), which is susceptible to disruption through interference, information in long-term working memory is not displaced if a task is interrupted. Long-term memory is not always reliable (see Chapter 8), though we know that experts demonstrate superior and more reliable memory in their domain of expertise. Experts appear to be able to develop and use retrieval cues in conventional working memory, and these cues greatly facilitate access to information in long-term memory. In short, experts appear able to expand their effective working memory capacity. Chess masters are able to perform quite astonishing feats, including playing blindfold chess. Originally it was thought that the superior performance of experts may have been due to experts acquiring and using larger chunks of knowledge stored in short-term memory, but examinations (e.g. Gobet, 1998) of the size of chunks expert chess players use indicated that chess masters must use much larger structures than the chunks Chase and Simon (1973a,b) had identified, and these structures must be stored in long-term memory. Chunking theory gave way to template theory (Gobet and Simon, 2000), in which templates are conceived as high-level schematic structures that may have evolved from perceptual chunks. For example, a chess template might be invoked by a player when a well-known position is encountered, such as the series of moves typically played in the opening known as 'Petrov Defence'. Template theory provides a good account of the ability of chess masters to play blindfold chess, as well as recall multiple chessboards presented in quick succession. Template theory, then, is entirely consistent with the notion of a long-term working memory.

Long-term working memory is consistent with accounts of expertise in a range of domains, an interesting example of which is acting. Actors must rapidly assimilate information about a role, and that information includes the lines the character speaks, the emotions that must be expressed at various points, the characters' intentions, and more. Noice and Noice (2006) note that actors have acquired with experience a vast amount of knowledge in long-term memory on human motivations, and intentions that are applied to those motivations. For example, an actor must understand the nature of alcoholic addiction in order to move across the stage in a performance, pick up a bottle of whisky, and say, 'This is how I solve my problems'. Understanding motivation and intention influence how the actor carries out this action and delivers the line. The intentions then serve as the retrieval cues that support the performance. Noice and Noice (2006) argue that this explains the ability of experts to go to a particular part of a performance (in rehearsal, for example) and to then play the part from there with conviction.

5.1.3 The role of general and specific methods

Schraagen (1993) carried out a more detailed examination of the problem-solving performance of different groups of experts and novices. Most studies of expertise have shown that experts draw upon a large body of domain knowledge when asked to solve a problem from their domain of expertise. Anderson's (1983) ACT* theory predicts that when domain knowledge is lacking, experts should fall back on general strategies, or 'weak methods'. Schraagen asked his participants to design an experiment in the area of sensory psychology. He compared the reasoning of domain experts (psychologists with around 10 years' experience in designing experiments in the area of sensory psychology) with 'design' experts (psychologists with around 10 years' experience in designing psychology experiments in general). The problem facing subjects was to design an experiment to investigate what people taste when they drink a given brand of cola. While domain knowledge was important (the domain experts generated better solutions), the form of the design experts' reasoning was comparable to that of the domain experts. When knowledge is lacking, it seems that there may be skills of intermediate generality that do transfer. These findings are at odds with theoretical frameworks that argue for the domain-specificity of expertise.

Schunn and Anderson (1999) carried out a similar study to examine whether expert scientists from different domains shared some skills. They asked domain experts (psychologists skilled in designing memory experiments), task experts (psychologists skilled in areas other than memory research), and undergraduate students studying different courses to think aloud while designing an experiment to investigate an unexplained aspect of memory. Analysis of protocols and performance data showed that the domain experts designed the best experiments. Domain experts and task experts differed in terms of domain-specific skills, while task experts and undergraduates differed on domain-general skills. Through the analysis of verbal protocols, the researchers were able to identify a much larger set of domain-general skills that are important in scientific reasoning.

5.2 Individual differences

The expert–novice paradigm contrasts the performance of experts and novices solving the same set of problems drawn from a given domain. Although the problems used are likely to be non-trivial to the novices, they scarcely present a problem in any meaningful sense to the experts. This is necessarily so because if the problems were truly challenging for the experts, novices would not be able to even begin to solve them. However, if the experts have not really been taxed with a 'problem', have we learned anything at all about expert problem solving? The expert–novice paradigm also tends to imply that novices know nothing, or know little of relevance. As we shall see, novices do not approach novel problems with 'empty heads'. They bring to bear whatever knowledge and strategies they are able to and, in doing so, it is clear that some novices are better learners than others. We shall now examine the extent to which implicit assumptions about the homogeneity of both novice and experts groups are reasonable.

5.2.1 Are all learners the same?

Novices have tended to be described in terms of what they do not have, or do not do. A more positive approach is to examine what novices *can* do, and ways in which they differ. In so doing, this work shifts the emphasis from problem-solving performance to learning and the acquisition of skill in problem solving. Models and theories of problem-solving performance must not only account for differences between skilled and less-skilled individuals; they must also explain how skill is acquired.

Some interesting work has examined differences between good and poor learners, and this has shed some light on what might mediate between practice and performance. Many of us will have noticed that people tend to differ in rate of learning. While it is an over-simplification to suggest that novices start with a blank slate, most novices begin from a position of not having much of the skill in question. If knowledge relevant to the skill does not mediate or support their performance early on, then what does?

Thorndyke and Stasz (1980) examined learning strategies, differentiating good from poor learners of map information. Good learners used more efficient techniques for encoding spatial information, more accurately determined what they knew and what they had yet to learn, and were better able to focus their attention on map elements they had not yet learned. Green and Gilhooly (1990) conducted a similar experiment, studying novices learning to use a statistical package on a mainframe computer. Good learners tended to adopt an exploratory approach to learning, made better use of worked examples from handouts, and evaluated their learning. Slower learners tended to over-use worked examples, generated and tested more erroneous hypotheses, and seemed to either ignore or fail to use error feedback. Both these studies suggest that good learners make effective use of metacognitive processes and strategies.

Chi *et al.* (1989) and Chi *et al.* (1994) have been especially interested in the role played by explanation in learning and, in particular, whether novices may be distinguished by the extent to which they generate explanations while solving problems. In their studies, they equated students for background knowledge (of physics and biology) and then analysed the think-aloud protocols students produced as they studied the problems. In one study, good learners seemed to spontaneously self-explain more than poor learners. Good learners used the examples they had studied to check their solutions, whereas poor learners used the examples to help them to find solutions. Chi *et al.* (1994) showed that prompting students to self-explain as they studied led to better problem solving than simply asking students to study the materials. Renkl (1997) showed that the self-explaining effect is not simply due to some students spending longer studying. In a study that controlled for time-on-task, Renkl found that quality of self-explanations reliably predicted learning success. Generating self-explanations, then, whether spontaneously or in response to a prompt to do so, seems to serve an elaborating role in early learning, aiding understanding and schema development. Schema development is of course central to skill development.

5.2.2 Can we enhance the rate of skill acquisition?

Sweller and his colleagues (Sweller, 1988; Sweller *et al.*, 1983) have demonstrated that schema acquisition can be retarded by the use of means–ends analysis. Paradoxically, the very strategy that novices appear to rely on in early learning (recall Section 4.1.2 and the study by Larkin *et al.*, 1980) has been shown to inhibit knowledge acquisition. Sweller *et al.* hypothesized that an emphasis on a goal (which occurs with the means–ends analysis strategy) might overload the system, leaving few resources available for inducing relevant schematic knowledge. They tested this hypothesis by de-emphasizing the goal in a set of kinematics (a branch of physics) problems given to one group of novices. For example, one problem ended in the following way: 'In 18 sec a racing car can start from rest and travel 305.1 m. Calculate the value of as many variables as you can'. A second group of novices received the same problems, but the final sentences were altered to include a specific goal, e.g. 'In 18 sec a racing car can start from rest and travel 305.1 m. What speed will it reach?'

Participants who were given the no-goal problems switched more swiftly to a working forwards strategy than did novices who were given the goal problems. One interpretation of these findings is that the presence of a goal biases individuals towards the use of a means–ends strategy, which imposes high processing demands. This would have the effect of reducing the available resources for acquiring knowledge about the relationships among principles. De-emphasizing the goal, then, could work by reducing working-memory load, thereby freeing resources. This would facilitate schema acquisition, thereby enhancing learning.

There is an alternative explanation for the facilitating effect of reduced goal specificity on learning. Vollmeyer *et al.* (1996) examined the effects of goal specificity and systematicity of learning strategy in learning and transfer within a complex dynamic system. Their findings were consistent with Sweller's claim that general problem-solving methods might enable a person to attain a specific goal, but do not promote learning of the overall structure of a problem space. Burns and Vollmeyer (2002) have taken this work further, and have shown that non-specific goals seem to aid learning by encouraging more hypothesis testing. Their work shows that it seems to be hypothesis testing, rather than the reduction in goal-specificity, that encourages learning.

There may be another dimension to the effects of goal specificity on learning and problem solving. Green (2002) argued that reducing goal specificity also has the effect of altering the way in which a problem comes to be represented. In her experiment, it was the nature of the problem representation that was crucial to performance, rather than the reduction in goal specificity. Green also points out that it is important to distinguish learning from problem solving. Instructions that led to swift learning seemed to result in poor problem solving, while instructions that seemed to lead to slower learning paid off later on by giving rise to better problem solving. Different instructions influence the nature of the task or problem representation, and this in turn affects both learning and problem-solving performance. This echoes the point we made earlier with regard to the impact of internal and external representations upon learning and problem solving.

Some recent studies by Haider and Frensch (1996, 1999a,b) have focused on ways in which we learn to ignore task-irrelevant information and process only task-relevant information. Haider and Frensch have shown that as we become more skilled, we typically learn to ignore redundant information. Not all individuals behave in the same way though. In one of their studies, they found that some individuals fail to reduce the amount of information heeded, even after extended practice. Green and Wright (2003) have extended these findings, examining what happens when two information sources associated with one event are presented. When individuals have a choice of information sources relevant to the task, they tend to prefer to use the first encountered source. Information reduction then serves to reduce processing of task-irrelevant, as well as duplicated (but possibly task-relevant) information. The assumption that we come to process less information is at odds with some theories of skill acquisition like ACT* and ACT-R (which you will explore in more detail in Chapter 19).

The studies we have discussed in this section provide some clues as to how individuals learn to solve problems more effectively. What is apparent is that learners do not all behave in the same way. Certain learning procedures and strategies facilitate knowledge acquisition, and there is evidence that problem representation again plays a key role.

Novices seem to differ from each other, then, and their rates of learning can vary. Do experts form a homogeneous group? We examine this question next.

5.2.3 Do experts differ?

It is sometimes tacitly assumed that experts form a homogeneous group, with considerable overlap among experts in what they know. If this is the case, then we may safely generalize from studies of experts, and talk about 'typical' expert problem-solving behaviour. However, it is likely that the assumption of homogeneity is at best an over-simplification.

This point has been made by Charness (1991) in his examination of chess skill, who found that chess masters do not know the full range of opening variations (there are some 50,000), middle-game combinations (around 1,817), and end-games (some 8,500). Indeed, it is questionable whether they could actually learn the full set. Instead, chess masters specialize in a subset of each of the three classes. Secondly, we can see that there are different kinds of expertise. The physics expert, for example, is an expert in a domain where knowledge of the principles themselves is sufficient to solve most of the problems that may be encountered. Chess masters exercise their skill in a domain where it is virtually impossible to learn all there is to know. There is not a body of 'principles' as such that are logically sufficient to solve most problems.

Chess, computer programming, and physics are, nonetheless, well-defined domains. In the case of chess in particular, certain problem states can become highly familiar, and stronger players can capitalize upon their ability to recognize good problem states. Sometimes, though, recognition hinders the construction of optimal representations.

Our notions of 'expertise' are becoming more sophisticated. Many of the examples considered here conform to the 'classic expert' model – the chess master, for example, or the physicist. Other kinds of expert are entrepreneurs, master chefs, and actors. A third category might be labelled specialists, and this would include individuals who have a niche of specialist knowledge or skill – a local historian, perhaps, or the person who is very knowledgeable on the works of a particular author or poet. The notion of 'expert' has a currency, and experts have status.

SUMMARY OF SECTION 5

- Neither experts nor novices form a homogeneous group.
- Skill in problem solving involves more than just the accumulation of knowledge.
- Problem-solving skill may be enhanced in a number of ways.

6 CONCLUSIONS

Nearly a century of research on problem solving has yielded some impressive findings. Important phenomena, such as insight and fixation, which have long taxed researchers, are now amenable to more rigorous, systematic investigation thanks to methodological and theoretical advances, not to mention the evolution of ever more sophisticated approaches to cognitive modelling. We now have a better understanding of analogical reasoning, helping us to appreciate how analogical reasoning occurs and why it sometimes fails. Advances have been made in understanding how we become skilled in solving problems from a wide range of complex domains, which have in turn led to a better understanding of expertise and learning.

We do not yet have a theory of problem solving; nor do we have a theory of learning, but progress is being made. Underpinning research on problem solving is a recurring theme, and it is this: representation is fundamental to problem solving, just as it is fundamental to many other areas of psychology. Problem representation is likely to be influenced by many variables, some of which we have only begun to explore. We hope we have stimulated your interest in this area of psychology sufficiently that you will see the potential for problem-solving research in terms of its wider application, as well as its theoretical significance.

FURTHER READING

Chronicle, E.P., MacGregor, J.N., and Ormerod, T.C. (2004) 'What makes an insight problem? The roles of heuristics, goal conception, and solution recoding in knowledge-lean problems', Journal of Experimental Psychology: Learning, Memory and Cognition, vol.30, pp.14–27. This paper provides an analysis of what is meant by the term 'insight problem' and four experiments investigating factors relevant to the solution of such problems.

Anderson, J.R. (2002) 'Spanning seven orders of magnitude: a challenge for cognitive modeling', Cognitive Science, vol.26, pp.85–112. In this paper the author applies his ideas to a range of problems, from those that are dealt with very rapidly to those that may require hours or weeks of thought.

REFERENCES

Adamson, R.E. and Taylor, D.W. (1954) 'Functional fixedness as related to elapsed time and set', *Journal of Experimental Psychology*, vol.47, pp.122–6.

Anderson, J.R. (1983) *The Architecture of Cognition*, Cambridge, MA, Harvard.

Anolli, L., Antonietti, A., Crisafulli, L., and Cantoia, M. (2001) 'Accessing source information in analogical problem-solving', *Quarterly Journal of Experimental Psychology*, vol.54A, pp.237–61.

Atwood, M.E. and Polson, P.G. (1976) 'A process model for water jug problems', *Cognitive Psychology*, vol.8, pp.191–216.

Blanchette, I. and Dunbar, K. (2000) 'How analogies are generated: the roles of structural and superficial similarity', *Memory and Cognition*, vol.28, pp.108–24.

Boshuizen, H.P.A. and Schmidt, H.G. (1992) 'On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices', *Cognitive Science*, vol.16, pp.153–84.

Burns, B.D. and Vollmeyer, R. (2002) 'Goal specificity effects on hypothesis testing in problem solving', *The Quarterly Journal of Experimental Psychology*, vol.55A, no.1, pp.241–61.

Card, S.K., Moran, T.P., and Newell, A. (1983) *The Psychology of Human–Computer Interaction*, Hillsdale, NJ, Erlbaum.

Charness, N. (1991) 'Expertise in chess: the balance between knowledge and search', in Ericsson, K.A. and Smith, J. (eds.) *Towards a General Theory of Expertise: Prospects and Limits*, Cambridge, MA, Cambridge University Press.

Chase, W.G. and Simon, H.A. (1973a) 'The mind's eye in chess', in Chase, W.G. (ed.) *Visual Information Processing*, New York, Academic Press.

Chase, W.G. and Simon, H.A. (1973b) 'Perception in chess', *Cognitive Psychology*, vol.4, pp.55–81.

Chi, M.T.H. (1997) 'Quantifying qualitative analyses of verbal data: a practical guide', *Journal of the Learning Sciences*, vol.6, pp.271–315.

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., and Glaser, R. (1989) 'Self-explanations: how students study and use examples in learning to solve problems', *Cognitive Science*, vol.13, pp.145–82.

Chi, M.T.H., de Leeuw, N., Chiu, M-H., and LaVancher, C. (1994) 'Eliciting self-explanations improves understanding', *Cognitive Science*, vol.18, no.3, pp.439–77.

Chi, M.T.H., Glaser, R., and Rees, E. (1982) 'Expertise in problem solving', in Sternberg, R.J. (ed.) *Advances in the Psychology of Human Intelligence*, Hillsdale, NJ, Erlbaum.

Davies, S.P. (2000) 'Move evaluation as a predictor and moderator of success in solutions to well-structured problems', *Quarterly Journal of Experimental Psychology*, vol.53A, no.4, pp.186–201.

Defeyter, M.A. and German, T.P. (2003) 'Acquiring an understanding of design: evidence from children's insight problem solving', *Cognition*, vol.89, pp.133–55.

De Groot, A.D. (1965) *Thought and Choice in Chess* (original edition in Dutch, 1946), The Hague, Mouton.

Dunbar, K. (2001) 'The analogical paradox: why analogy is so easy in naturalistic settings, yet so difficult in the psychology laboratory', in Gentner, D., Holyoak, K.J., and Kokinov, B. (eds.) *Analogy: Perspectives from Cognitive Science*, Cambridge, MA, MIT Press.

Duncker, K. (1945) 'On problem solving', *Psychological Monographs*, vol.58, whole, no.270, pp.1–113.

Ericsson, K.A. (1991) 'Prospects and limits of the empirical study of expertise: an introduction', in Ericsson, K.A. and Smith, J. (eds.) *Towards a General Theory of Expertise: Prospects and Limits*, Cambridge, MA, Cambridge University Press.

Ericsson, K.A. and Harris, M. (1990) 'Expert chess memory without chess knowledge. A training study', poster presentation at the 31st Meeting of the Psychonomics Society, New Orleans.

Ericsson, K.A. and Kintsch, W. (1995) 'Long-term working memory', *Psychological Review*, vol.102, no.2, pp.211–45.

Ericsson, K.A. and Polson, P.G. (1988) 'Memory for restaurant orders', in Chi, M.T.H., Glaser, R., and Farr, M. (eds.) *The Nature of Expertise*, Hillsdale, NJ, Erlbaum.

Ericsson, K.A., Krampe, R.T., and Tesch-Rohmer, C. (1993) 'The role of deliberate practice', *Psychological Review*, vol.100, no.3, pp.363–406.

Falkenhainer, B., Forbus, K.D., and Gentner, D. (1986) 'The structure-mapping engine', *Proceedings of the Meeting of the American Association for Artificial Intelligence*, pp.272–7.

Gentner, D. (1983) 'Structure-mapping: a theoretical framework for analogy', *Cognitive Science*, vol.7, pp.155–70.

Gentner, D. (2003) 'Why we're so smart', in Gentner, D. and Goldin-Meadow, S. (eds.) *Language in Mind: Advances in the Study of Language and Thought*, Cambridge, MA, MIT Press.

Gentner, D. and Gentner, D.R. (1983) 'Flowing waters or teeming crowds: mental models of electricity', in Gentner, D. and Stevens, A.L. (eds.) *Mental Models*, Hillsdale, NJ, Lawrence Erlbaum Associates. (Reprinted in Brosnan, M.J. (ed.) *Cognitive Functions: Classic Readings in Representation and Reasoning*, Eltham, London, Greenwich University Press.)

Gentner, D. and Markman, A.B. (1997) 'Structure mapping in analogy and similarity', *American Psychologist*, vol.52, no.1, pp.45–56.

Gentner, D., Bowdle, B., Wolff, P., and Boronat, C. (2001) 'Metaphor is like analogy', in Gentner, D., Holyoak, K.J., and Kokinov, B.N. (eds.) *The Analogical Mind: Perspectives from Cognitive Science*, Cambridge, MA, MIT Press.

German, T.P. and Barrett, H.C. (2005) 'Functional fixedness in a technologically sparse culture', *Psychological Science*, vol.16, pp.1–5.

Gick, M.L. and Holyoak, K.J. (1980) 'Analogical problem solving', *Cognitive Psychology*, vol.12, pp.306–55.

Gilhooly, K.J., McGeorge, P., Hunter, J., Rawles, J.M., Kirby, I.K., Green, C., and Wynn, V. (1997) 'Biomedical knowledge in diagnostic thinking: the case of electrocardiogram (ECG) interpretation', *European Journal of Cognitive Psychology*, vol.9, no.2, pp.199–223.

Gilhooly, K.J., Phillips, L.H., Wynn, V., Logie, R.H., and Della Sala, S. (1999) 'Planning processes and age in the 5 disk Tower of London task', *Thinking and Reasoning*, vol.5, no.4, pp.339–61.

Glucksberg, S. and Danks, J.H. (1968) 'Effects of discriminative labels', *Journal of Verbal Learning and Verbal Behaviour*, vol.7, pp.72–6.

Glucksberg, S., Gildea, P., and Bookin, H. (1982) 'On understanding non-literal speech: can people ignore metaphors?' *Journal of Verbal Learning and Verbal Behavior*, vol.1, pp.85–96.

Gobet, F. (1998) 'Expert memory: a comparison of four theories', *Cognition*, vol.66, pp.115–52.

Gobet, F. and Simon, H.A. (2000) 'Five seconds or sixty? Presentation time in expert memory', *Cognitive Science*, vol.24, pp.651–82.

Green, A.J.K. (2002) 'Learning procedures and goal specificity in learning and problem solving tasks', *European Journal of Cognitive Psychology*, vol.14, no.1, pp.105–26.

Green, A.J.K. and Gilhooly, K.J. (1990) 'Individual differences and effective learning procedures: the case of statistical computing', *International Journal of Man-Machine Studies*, vol.33, pp.97–119.

Green, A.J.K. and Wright, M.J. (2003) 'Reduction of task-relevant information in skill acquisition', *European Journal of Cognitive Psychology*, vol.15, no.2, pp.267–90.

Haider, H. and Frensch, P.A. (1996) 'The role of information reduction in skill acquisition', *Cognitive Psychology*, vol.30, no.3, pp.304–37.

Haider, H. and Frensch, P.A. (1999a) 'Eye movement during skill acquisition: more evidence for the information reduction hypothesis', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.25, no.1, pp.172–90.

Haider, H. and Frensch, P.A. (1999b) 'Information reduction during skill acquisition: the influence of task instruction', *Journal of Experimental Psychology: Applied*, vol.5, no.2, pp.129–51.

Hall, T. (1970) *Carl Friedrich Gauss: A Biography* (translated by A. Froderberg), Cambridge, MA, MIT Press.

Holding, D.H. (1985) *The Psychology of Chess Skill*, Hillsdale, NJ, Erlbaum.

Holding, D.H. and Reynolds, R.I. (1982) 'Recall or evaluation of chess positions as determinants of chess skill', *Memory and Cognition*, vol.10, pp.237–42.

Holyoak, K.J. and Koh, K. (1987) 'Surface and structural similarity in analogical transfer', *Memory and Cognition*, vol.15, pp.332–40.

Holyoak, K.J. and Thagard, P. (1989) 'A computational model of analogical problem solving', in Vosniadou, S. and A. Ortony (eds.) *Similarity and Analogical Reasoning*, Cambridge, MA, Cambridge University Press.

Jeffries, R., Polson, P.G., Razran, L., and Atwood, M.E. (1977) 'A process model for missionaries-cannibals and other river crossing problems', *Cognitive Psychology*, vol.9, pp.412–20.

Jones, G. (2003) 'Testing two theories of insight', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.29, pp.1017–27.

Keane, M. (1988) *Analogical Problem Solving*, Chichester, Ellis Horwood.

Keane, M.T. (1994) 'Constraints on analogical mapping: a comparison of three models', *Cognitive Science*, vol.18, no.3, pp.387–438.

Knoblich, G., Ohlsson, S., Haider, H., and Rhenius, D. (1999) 'Constraint relaxation and chunk decomposition in insight problem solving', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.25, no.6, pp.1543–55.

Kolers, P.A. (1975) 'Memorial consequences of automatized encoding', *Journal of Experimental Psychology: Human Learning and Memory*, vol.1, pp.689–701.

Larkin, J., McDermott, J., Simon, D.P., and Simon, H.A. (1980) 'Models of competence in solving physics problems', *Cognitive Science*, vol.4, pp.317–45.

Lesgold, A.M., Rubinson, H., Feltovich, P.J., Glaser, R., Klopfer, D., and Wang, Y. (1998) 'Expertise in a complex skill: diagnosing X-ray pictures', in Chi, M.T.H., Glaser, R., and Farr, M. (eds.) *The Nature of Expertise*, Hillsdale, NJ, Erlbaum.

Logan, G.D. (1988) 'Toward an instance theory of automatization', *Psychological Review*, vol.95, pp.492–527.

Luchins, A.S. and Luchins, E.H. (1959) *Rigidity of Behaviour*, Eugene, OR, University of Oregon Press.

Luger, G.F. (1976) 'The use of state-space to record the behavioural effects of subproblems and symmetries on the Tower of Hanoi problem', *International Journal of Man-Machine Studies*, vol.8, pp.411–21.

Lung, C-T. and Dominowski, R.L. (1985) 'Effects of strategy instructions and practice on nine-dot problem solving', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.11, pp.804–11.

MacGregor, J.N., Ormerod, T.C., and Chronicle, E.P. (2001) 'Information-processing and insight: a process model of performance on the nine-dot and related problems', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.27, pp.176–201.

McKeithen, K.B., Reitman, J.S., Rueter, H.H., and Hirtle, S.C. (1981) 'Knowledge organization and skill differences in computer programmers', *Cognitive Psychology*, vol.13, pp.307–25.

Neisser, U., Novick, R., and Lazar, R. (1963) 'Searching for ten targets simultaneously', *Perceptual and Motor Skills*, vol.17, pp.955–61.

Noice, H. and Noice, T. (2006) 'Artistic performance: acting, ballet and contemporary dance', in Ericsson, K.A., Charness, N., Feltovich, P.J. and Hoffman, R.R. (eds.) *The Cambridge Handbook of Expertise and Expert Performance*, Cambridge, Cambridge University Press.

Ohlsson, S. (1992) 'Information processing explanations of insight and related phenomena', in Keane, M.T. and Gilhooly, K.J. (eds.) *Advances in the Psychology of Thinking*, London, Harvester Wheatsheaf.

Ormerod, T.C., MacGregor, J.N., and Chronicle, E.P. (2002) 'Dynamics and constraints in insight problem solving', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.28, no.4, pp.791–9.

Rattermann, M.J. and Gentner, D. (1998) 'More evidence for a relational shift in the development of analogy: children's performance on a causal-mapping task', *Cognitive Development*, vol.13, pp.453–78.

Renkl, A. (1997) 'Learning from worked-out examples: a study of individual differences', *Cognitive Science*, vol.21, no.1, pp.1–29.

Schoenfeld, A.H. and Herrmann, D.J. (1982) 'Problem perception and knowledge structure in expert and novice mathematical problem solvers', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.8, no.5, pp.484–94.

Schraagen, J.M. (1993) 'How experts solve a novel problem in experimental design', *Cognitive Science*, vol.17, no.2, pp.285–309.

Schultetus, S. and Charness, N. (1999) 'Recall vs. position evaluation revisited: the importance of position-specific memory in chess skill', *American Journal of Psychology*, vol.112, no.4, pp.555–69.

Schunn, C.D. and Anderson, J.R. (1999) 'The generality/specificity of expertise in scientific reasoning', *Cognitive Science*, vol.23, no.3, pp.337–70.

Simon, H.A. and Chase, W.G. (1973) 'Skill in chess', *American Scientist*, vol.61, pp.394–403.

Simon, H.A. and Hayes, J.R. (1976) 'The understanding process: problem isomorphs', *Cognitive Psychology*, vol.8, pp.165–90.

Simon, H.A. and Reed, S.K. (1976) 'Modelling strategy shifts on a problem solving task', *Cognitive Psychology*, vol.8, pp.86–97.

Snoddy, G.S. (1926) 'Learning and stability', *Journal of Applied Psychology*, vol.10, pp.1–36.

Spellman, B.A. and Holyoak, K.J. (1992) 'If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles', *Journal of Personality and Social Psychology*, vol.62, pp.913–33.

Sternberg, R.J. (1995) 'Expertise in complex problem solving', in Frensch, P.A. and Funke, J. (eds.) *Complex Problem Solving*, Hillsdale, NJ, Erlbaum.

Sweller, J. (1988) 'Cognitive load during problem solving: effects on learning', *Cognitive Science*, vol.12, no.2, pp.257–85.

Sweller, J., Mawer, R.F., and Ward, M.R. (1983) 'Development of expertise in mathematical problem solving', *Journal of Experimental Psychology: General*, vol.112, no.4, pp.639–61.

Thomas, J.C. Jr. (1974) 'An analysis of behaviour in the hobbits–orcs problem', *Cognitive Psychology*, vol.6, pp.257–69.

Thorndyke, P.W. and Stasz, C. (1980) 'Individual differences in procedures for knowledge acquisition from maps', *Cognitive Psychology*, vol.12, pp.137–75.

Vollmeyer, R., Burns, B.D., and Holyoak, K.J. (1996) 'The impact of goal specificity on strategy use and the acquisition of problem structure', *Cognitive Science*, vol.20, pp.75–100.

Voss, J.F., Greene, T.R., Post, T.A., and Penner, B.C. (1983) 'Problem solving skill in the social sciences', in Bower, G. (ed.) *The Psychology of Learning and Motivation*, vol.17, New York, Academic Press.

Weisberg, R.W. and Alba, J.W. (1981) 'An examination of the alleged role of "fixation" in the solution of several "insight" problems', *Journal of Experimental Psychology: General*, vol.110, pp.169–92.

Wolff, P. and Gentner, D. (2000) 'Evidence for role-neutral initial processing of metaphors', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.26, pp.529–41.

Zhang, J. and Norman, D.A. (1994) 'Representations in distributed cognitive tasks', *Cognitive Science*, vol.18, no.1, pp.87–122.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

CHAPTER 11

JUDGEMENT AND DECISION MAKING

Peter Ayton

1 INTRODUCTION

How do people make judgements and decisions? The question has become a steadily increasing preoccupation of cognitive psychology. Plainly, making decisions is a fundamental and everyday human and animal (and perhaps machine) activity. Yet, until the 1950s psychologists had hardly given the question any serious thought. Doubtless this had something to do with the dominance of the behaviourist school of thought throughout the first half of the twentieth century. The behaviourists assumed that human behaviour could be explained entirely in terms of reflexes, stimulus–response associations, and the effects of reinforcers upon them. Accordingly, they shunned the study of mental processes and entirely excluded 'mental' terms like desires and goals. As a historical consequence, the foundations of decision research, and hence its contemporary shape, have been strongly influenced by thinking from disciplines *outside* psychology – specifically from mathematics and economics.

This influence from outside psychology left its mark – mathematicians and economists have different concerns to psychologists. The question posed and pursued by thinkers from outside psychology was not how *do* people actually make decisions but how, ideally, *should* decisions be made? What are *good* judgements and decisions and how should we recognize them? As we will see, *behavioural* judgement and decision research – the investigation of *how* people make decisions – has been strongly influenced by a fundamental underlying premise: that the objective of decision making should be to make the 'best' choice, and that the best choice can, by some method, be computed.

Judgement and decision making are sometimes distinguished on the basis that *judgements are what underlie decisions*. Judgements can be estimates of some objective quantity – how far away is this object? How dangerous is that hobby? Decisions typically reflect judgements of the qualities of options – but also the preferences of the decision maker.

Of course, real people are not idealized decision-making machines or supercomputers; they do not have unlimited time, knowledge, and computational power but a rather limited information-processing capacity. Accordingly, not infrequently, people make mistakes – for example, they may overlook or forget important considerations; they also get bored, suffer anxiety, and may not always be sure quite what they want or are trying to achieve. As a consequence, what people do is not always quite the same as what they themselves would agree that they *should* do.

1.1 Theories of decision making

Psychologists are of course interested in understanding what people actually do, but this has very often been studied in comparison to what it has been assumed they *should* do. As a result, there are two types of theory of decision making – the 'ought' and the 'is' – commonly referred to as **normative** and **descriptive** theories respectively. Normative theories define the supposed ideal decision while descriptive theories attempt to characterize how people actually make decisions.

The very existence of this dichotomy suggests that perhaps human decision making is faulty. Indeed, debating whether or not people are essentially rational or irrational decision makers has long been a preoccupation of researchers in this field (see Cohen, 1981), just as the rationality of thought has been a key concern for researchers studying human reasoning (as you will see in Chapter 12). However, noting a disparity between the ideal and the actual should not, in itself, cause us to leap to the conclusion that there is something *fundamentally* wrong with the way people make judgements and decisions. In other areas of cognitive psychology, such a step would be seen as clearly absurd; for instance, human memory is manifestly fallible and yet we do not conclude from this that people's memories are inherently inadequate for the purpose of living their lives.

While persistent errors of judgement or choice could be taken to indicate a fundamental irrationality, researchers in judgement and decision making have tended to adopt a similar position to researchers working in vision. For instance, vision scientists do not conclude from the robustness of the Müller–Lyer illusion (see Chapter 3, Section 1) that people are generally poor at inferring object lengths – let alone that visual perception is fundamentally incompetent. Nevertheless, as we will see, people do make judgements and decisions that are inconsistent with normative theory.

1.2 Supporting decision making

If people don't behave as normative theories prescribe, what can be done about it? What should people do to make better choices? What instruction, modes of thinking, or decision aids can help real people to make better decisions? A third strand of research straddling the normative and descriptive – the **prescriptive** approach – investigates how to help people make better decisions. One prescriptive approach is decision analysis. Decision analysis is the attempt to help people to make better decisions that conform to normative theory. However, decision analysis is more than just that; helping people to understand and explicate their own objectives and values, search for options and evidence, and appreciate their implications is not a straightforward matter. Decision analysis uses a number of techniques, including decision trees (which you will meet in Section 2.1), to help people decompose complex decisions into more manageable components, elicit values and beliefs for the elements, and apply normative principles to their reintegration.

SUMMARY OF SECTION 1

- Judgements underlie decisions.
- Researchers distinguish between actual (descriptive) and ideal (normative) decision making.
- Decision analysis can support people in making better decisions – a prescriptive approach.

2 NORMATIVE THEORY OF CHOICE UNDER RISK

In many situations where we must make a choice, we will be uncertain about whether the possible outcomes will turn out to be good or not so good. Consequently, risk is an inescapable fact of life. Some sorts of risky decisions are easy to imagine: a person may have to consider whether to continue living with a debilitating health condition or risk surgery that might help but could leave them worse off. Investment decisions often involve contemplating whether to put money in a safe investment with a small return or a riskier investment that might yield a lot of money but could lose everything. Decisions of this sort can be analysed as gambles. Gambling is the dominant metaphor in decision research as gambles involve *uncertainty* about what will happen.

The most extensively applied normative model of risky choice is called subjective expected utility theory (SEU). This theory is an extension by Savage (1954) of the 'expected utility' theory published by von Neumann and Morgenstern (1944) in their book *Theory of Games and Economic Behavior*. Von Neumann and Morgenstern's analysis was applied to games of chance with known or computable probabilities. Savage's extension of the theory allowed for what he called 'personal probabilities' – commonly referred to nowadays as 'subjective probabilities'. Savage's generalization of the theory allows it to be applied to decision situations where no objective mathematical probabilities are available and where judgements may be no more than expressed beliefs about likelihoods.

For example: imagine contemplating an invitation to a picnic. Suppose you have to write an essay over the weekend in question but do not want to miss out on anything really good. On the other hand, you wouldn't want to waste time at a boring or horrible picnic. You may be unsure whether it will rain or not; whether Tarquin (an individual about whom you have very strong views) will be present or not; and there could be any number of other factors that would affect the value of accepting the invitation. So how should you decide?

According to standard normative theory, a rational decision maker should trade off the value of all the possible outcomes by the likelihood of obtaining them. Just as the value of a lottery ticket will vary according to both the value of the prizes *and* the chances of winning them, so, according to the normative theory of choice, do choice options. What is the value of an idyllic or dreadful picnic? What is the value of any of the alternative activities you could indulge in? How likely is it to rain or that Tarquin will be there? All the relevant elements must be quantified and combined to compute the optimal decision. How this is done is illustrated in the next section where the technique of decision analysis is described.

2.1 Prescriptive application of normative theory: decision analysis

Decision analysis is a technology based on SEU that was developed in the 1960s to improve decision making (see Raiffa, 1968; Schlaifer, 1969). Decision analysts use normative theory to represent decision problems so that, ideally at least, the normatively correct decision can be computed. In the classic decision analytic framework (von Winterfeldt and Edwards, 1986) numerical probabilities are assigned to all the different events identified in a decision tree. The decision tree is simply a means of representing or modelling the decision. The best alternative is then selected by combining the probabilities and the utilities corresponding to the possible outcomes associated with each of the possible alternatives.

Figure 11.1 shows a simple decision tree for our student trying to decide whether to go on a picnic or stay at home and write an essay. The tree portrays two possible actions or events – picnic or essay – and three possible future events (weather conditions) that would

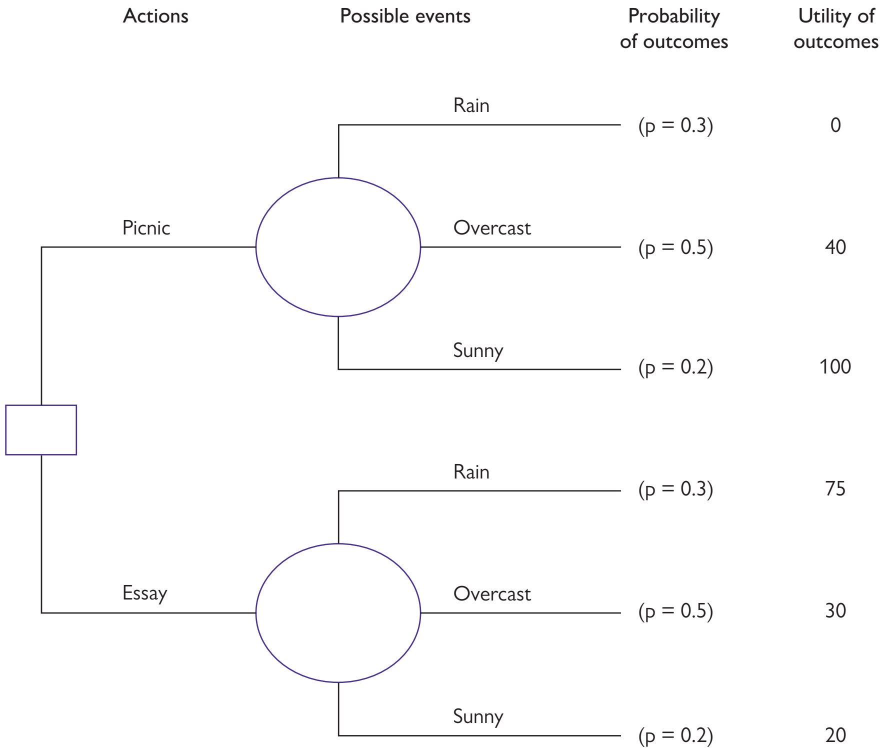


FIGURE 11.1 Decision tree for deciding whether to go on a picnic or write an essay.

affect the value or utility of the resulting six identifiable outcomes. Of course, there could be many more options (students' weekend options involve more than essays and picnics) and possible future conditions (there may well be more than the weather to consider). The utilities, as in this case, will often reflect subjective evaluations of the quality of the outcomes – though for business or financial decisions they might reflect monetary profits. Here, the scale for utility is arbitrary – let's suppose the student was asked to rate each of the six possible outcomes on a 0–100 scale, assigning 0 to the worst outcome, 100 to the best, and scaling the others relative to those. A picnic in the sun is rated the best outcome and a picnic in the rain the worst. Writing an essay at home is affected by the weather, albeit differently, because, as any student knows, a sunny day is the worst time to have to stay in and work – especially if you know your friends are out having fun. These figures might not reflect your own utilities for these outcomes – utility is personal. You might revel in rainy picnics – if so, your utilities for this decision would be quite different.

We also need likelihoods for the three different weather conditions. In some countries weather forecasters routinely broadcast these, but as long as they accurately reflect our real beliefs, we could use our own judgements. Decision theory can only tell you how to decide given your beliefs about the utility and likelihood of the relevant events. With these data, we can now compute the expected utility of the two actions identified in the tree. The expected utility of each *outcome* is defined as the probability or likelihood of that outcome (P) multiplied by its utility (U). The expected utility of an *action* is the sum of such expected utilities for each of the possible *outcomes*. So, the expected utility of the picnic *action* is the sum of the expected utilities for the rainy picnic, the overcast picnic, and the sunny picnic:

$$\begin{aligned} SEU(picnic) = \Sigma P \times U &= P(rainy\ picnic) \times U(rainy\ picnic) \\ &\quad + P(overcast\ picnic) \times U(overcast\ picnic) \\ &\quad + P(sunny\ picnic) \times U(sunny\ picnic) \\ &= (0.3 \times 0) + (0.5 \times 40) + (0.2 \times 100) \\ &= 40 \end{aligned}$$

That is, by multiplying the utility of each outcome by the likelihood of obtaining it, we can calculate that the expected utility for the picnic is 40. Similarly, we can calculate that the expected utility for the essay is 41.5. Because the expected utility of the essay is greater than that for the picnic, *this* student should write the essay.

The difference between 41.5 and 40 may seem rather close, but remember that the numerical scales are arbitrary. In practice, a decision analyst using this procedure as part of the *prescriptive* approach to assist decision making would go back and check that reasonable variations in the values entered in the tree do not alter the decision. If they do, then the user must be sure that the numbers genuinely reflect their beliefs and values.

Since its initial inception, the role of decision analysis has changed (Phillips, 1989). Nowadays, decision analysts view decision trees as tools to aid thinking, not as providing solutions (von Winterfeldt and Edwards, 1986). The theory of **requisite decision modelling** (Phillips, 1984) claims that models of decisions need only be sufficient in structure and content to resolve the issues at hand. A good model captures the essential elements of a decision situation for the purposes of the decisions to be made. An iterative procedure is followed, involving constructing the model, analysis, model refinement, and subsequent re-analysis. At the point when no additional intuitions emerge from further analysis, the model is said to be *requisite*. The claim is that this procedure helps to develop a shared understanding and fosters commitment to the way forward. Note that these are social purposes. The technical computations of a decision analysis are less than half the story of improving decision making. Constructing representations of decisions and eliciting values are not achieved by mechanical operations; they often involve deep thought – or what some (see Watson and Buede, 1987) call 'soul searching'. Perhaps the greatest virtue of decision analysis is that it obliges decision makers to make explicit all the bases of a decision.

In the next section, we shall examine some of the principles underlying SEU, and consider the extent to which SEU accurately describes human decision making.

✱ ACTIVITY 11.1

Try to produce a decision tree for two activities you might consider doing. For example, you might want to compare going to a party with going to the cinema. Your enjoyment of a party may depend on who else is going and where the party is. Going to the cinema may be fun if you can see a film you particularly want to see, or might be just an opportunity to while away some time if you are indifferent to the films on show. Think about which you would choose to do. Now try computing the expected utility for the two actions. Which action does the analysis suggest you *should* choose? Are they the same and, if not, why do you think they differ?

COMMENT

The calculations illustrate how the expected utility theory defines a normative decision, but why follow this procedure? The assumption is that you might not make such a good decision by relying on unaided intuition. The decision tree can help you to clarify the relevant events and the structure of the decision, while the computations based on your stated values will follow the normative theory. Alas, there is no normative technique for eliciting the correct structure of decisions from individuals. However, in decision-analytic practice decision trees are used more to facilitate thought than to definitively represent complex decisions; a realistic portrayal of all relevant act–event combinations might result in a complicated mess.

2.2 Axioms underlying subjective expected utility theory

Mathematicians such as von Neumann and Savage established that SEU is implied by the acceptance of certain principles or axioms: **comparability**, **transitivity**, **dominance**, **independence**, and **invariance**. According to SEU, if a decision maker violates one or more of these axioms, then their choices will not maximize expected utility and so will not be normative. The axioms therefore define a kind of coherence to our choices and give them internal consistency. We shall look at them in more detail now:

- Comparability (or completeness): If you have to evaluate two alternatives, A and B, you must be able to say whether you:
    1. prefer A to B, or
    2. prefer B to A, or
    3. are indifferent between A and B.
- Transitivity: If you prefer A to B and B to C, then you must prefer A to C. That is, choices should be capable of being ordered.
- Dominance: An option is dominant and must be preferred if, when compared to another option, it is better in at least one respect and at least as good or better in every other respect. *Dominated* options must never be preferred.
- Independence: If there is some outcome that is unaffected by, or independent of, your choice, then this outcome should not affect your choice.
- Invariance: Different representations of the same choice problem should result in the same choices. That is, the preference for options should be independent of how they are described.

2.3 Violations of the axioms

Although the axioms might strike you as uncontentious and straightforward, they can be questioned. For example, the comparability axiom is threatened by claims that people may not be indifferent just because they are unable to say which of two states they prefer. Curiously, even the original architects of the theory admitted that it might not always be a reasonable assumption:

> *We have conceded that one may doubt whether a person can always decide which of two alternatives … he prefers. If the general comparability assumption is not made, a mathematical theory … is still possible.*
>
> (von Neumann and Morgenstern, 1944, pp.19–20)

The axioms also seem to vary somewhat in their intuitive appeal; while independence and transitivity might not be obvious requirements for rational choice, dominance and invariance appear essential. Nonetheless, psychologists have shown that, under certain conditions, systematic violations of each of the axioms can be observed in people's choices (Tversky and Kahneman, 1986). Since violations of the axioms imply that people are not choosing according to the normative theory, we could conclude one of two things: either that there is something wrong with the choices or that there is something wrong with the normative theory (or perhaps both). In any case, as we shall now see, SEU does not provide a good description of actual human choices.

2.3.1 Violations of transitivity

Observed violations of the transitivity axiom have generally led to the conclusion that people's choices are not ideal. For example, Tversky (1969) asked people to state their preferences between pairs of college applicants rated on three dimensions – intelligence, emotional stability, and social facility (as in Table 11.1).

Now try Activity 11.2.

If people reliably mapped all the dimension scores of each option onto a common currency of utility, then *systematic* violations of transitivity would not occur – so demonstrations of intransitive preference are revealing about the nature of the choice process. For Tversky (1969, p.46), this was key: 'The main

✱ ACTIVITY 11.2

Consider the following pairs of applicants in Table 11.1: A–B, B–C, C–D, D–E, and E–A. For each pair, write down which of the two applicants you would prefer given their ratings. Do note, however, that you should weight intelligence more highly than either of the other dimensions.

COMMENT

Tversky's subjects were presented with all possible pairs of applicants (together with some others), one pair at a time, and were similarly told to weight intelligence more than the other two dimensions. Subjects typically preferred A to B; B to C; C to D; and D to E. However, violations of transitivity were demonstrated by the typical simultaneous preference for E over A. Did your own preferences coincide with these? If not, try to work out why Tversky's subjects might have adopted the preferences they did.

TABLE 11.1 Ratings of five applicants on three dimensions

Applicant	Intelligence	Emotional stability	Social facility
A	69	84	75
B	72	78	65
C	75	72	55
D	78	66	45
E	81	60	35

interest in the present results lies not so much in the fact that transitivity can be violated but rather in what these violations reveal about the choice mechanism.' Tversky suggested ways in which decision making might be rendered less cognitively demanding by applying decision rules that simplify the task. He offered two hypotheses about the choice process: (1) people compare the alternatives on each dimension in turn, rather than *evaluating* each option on all dimensions before comparing overall evaluations; and (2) people ignore dimensions on which the alternatives – even if discriminable – are rated similarly.

For example, when comparing successive pairs in the chain (such as A and B) on intelligence, subjects may decide that the difference between them is negligible – and so, in the interests of simplifying the decision, ignore it altogether. However, small differences add up – at the ends of the chain the difference in intelligence between A and E is too big to ignore – hence, the observed pattern of intransitivity. Note that this explanation (that people try to simplify decisions by ignoring information) assumes that people have limited information-processing capacity.

In relation to SEU, intransitivity is an irrational pattern of choice, but it may be reassuring to note that it is not a uniquely human condition. For instance, in an experiment where bees chose between artificial flowers that offered varying amounts of nectar with varying degrees of accessibility, Shafir (1994) found that they violated transitivity in their foraging preferences for flowers. As bees have been successfully foraging for millions of years it is tempting to assume that perhaps the costs – intransitive preferences cannot maximize expected utility – are outweighed by the gains – presumably, reduced information processing.

When confronted with evidence of intransitivity in their choices, people typically immediately concede that there is some inconsistency and are usually willing to change their choices to preserve transitivity. Hence, they seem to endorse the normative status of the axiom even though their violations show that transitivity is not descriptive of human choice.

2.3.2 Violations of the independence axiom

Violations of the independence axiom are more problematic, and have proved a serious challenge to both the normative and descriptive status of SEU. The first challenge came from the French economist and Nobel laureate Maurice Allais, who published a paper in 1953 describing what is now called the Allais paradox (Allais, 1953; 1979).

Allais observed that people are reluctant to exchange a certain prospect of something wonderful (e.g. receiving $1,000,000) for a not quite certain prospect of something even more wonderful (e.g. 99 per cent chance of receiving $5,000,000). The paradox occurs because if both the above prospects are reduced in likelihood by a similar amount (so that neither offers certainty) people *are* usually willing to exchange a smidgen of likelihood for a substantial increase in benefit. Box 11.1 shows how this is a problem for SEU.

Allais made it perfectly clear that he considered that the intuitions that produced the paradox *should* overrule the independence axiom; that is, the normative theory was not valid. He even claimed: 'It is quite disappointing to have to exert so much effort to prove the illusory character of a formulation whose over-simplification is evident to anyone with a little psychological intuition' (Allais and Hagen, 1979, p.105). Others, including Savage (who, embarrassingly, initially succumbed to the paradox in his own choices), felt differently and argued that the intuitions underlying the choices were wrong and that the theory was normatively correct.

When made aware that they are violating the independence axiom, people sometimes alter their choices to conform to it (Keller, 1985) but sometimes – even after a thorough explanation of its virtues – they don't (Slovic and Tversky, 1974). Slovic and Tversky suggested that people may alter their choices to concur with the axiom not through appreciating the merits of doing so, but because they might be intimidated by the suggestion that not doing so would be irrational. Their paper concludes with a delightful imaginary debate between Savage and Allais wherein Savage insists that people only reject the axiom when they do not understand it, while Allais (who plausibly claimed to both understand and reject the axiom) asks how Savage could distinguish between failure to understand the axiom and enlightened rejection of it. The debate highlights an irresolvable conflict between two

BOX 11.1 The Allais paradox

Table 11.2 shows two separate situations where you can choose to take part in one of two lotteries and draw one ticket from the lottery you choose. In Situation 1 you can choose between lotteries A and B. If you are like most people you would choose A, as this guarantees $1,000,000. B could deliver $5,000,000 but there is a small chance of ending up with nothing at all.

However, when faced with the choice in Situation 2, between C and D, most people prefer D – now they are willing to face a very slight increase in the prospect of getting nothing at all in order to have a chance of winning $5,000,000. To see how this violates the independence axiom, simply cover up the last column – now the two situations appear identical. As the contents of the last column are identical for A and B in Situation 1 (i.e. $1,000,000), and also for C and D in Situation 2 ($0), then, according to the axiom, the information in this column should not influence your choice. So, if you prefer A to B, you should also prefer C to D.

TABLE 11.2 The Allais paradox as a choice of lotteries: each lottery involves 100 tickets. The table shows the number of tickets that win anything from $0 to $5,000,000

		Lottery ticket numbers (1–100)		
		1	2–11	12–100
Situation 1	Choice A	$1,000,000	$1,000,000	$1,000,000
	Choice B	$0	$5,000,000	$1,000,000
Situation 2	Choice C	$1,000,000	$1,000,000	$0
	Choice D	$0	$5,000,000	$0

different intuitions – those that support the axiom and those that support the pattern of choices in the Allais paradox. Ultimately, rather like the Ten Commandments (which are also often violated), the normative status of SEU and its axioms is not in any sense a demonstrable truth – they only appeal (or not) as principles to live by.

SUMMARY OF SECTION 2

- SEU provides a normative theory of decision making under uncertainty.
- Decision analysis offers a prescription for making decisions using SEU.
- Conforming to SEU is equivalent to adhering to certain axioms.
- Human decision making has been shown to violate these axioms, implying that it is not adequately described by SEU.

3 FINDINGS FROM BEHAVIOURAL DECISION RESEARCH

Most decision researchers accept the normative status of SEU but also consider that it does not describe human decision making. Some 30 years after the emergence of SEU, Slovic *et al.* (1977, p.9) reviewed the psychological literature and commented: '… during the past 5 years, the proponents of SEU have been greatly outnumbered by its critics'. Edwards (1992) polled an all-star cast of leading decision theorists at a conference. They unanimously endorsed traditional SEU as the appropriate normative model but unanimously agreed that people don't behave as the model requires. Nonetheless, and perhaps in spite of the survival of SEU as a normative theory (albeit on the basis of opinion polls), Allais was awarded the Nobel Prize for Economics in 1988.

Violations of the axioms of SEU imply that it does not provide a valid description of human decision making. There is now a considerable mass of empirical evidence indicating that SEU does not predict human decisions either. One piece of evidence comes from Edwards (1955), who offered experimental subjects choices between bets of equal expected value such as the choice between Gambles A and B in Figure 11.2 below. If you accept Gamble A, it will give a 0.6 probability (or 60 per cent chance) of winning £2.00, and a 0.4 chance of £4.00; Gamble B gives a 0.2 probability of winning £14.00 and a 0.8 probability of winning nothing.

If we assume that the amounts of money are in direct proportion to people's utility for each outcome, then we can calculate the expected utility (EU) for the bets:

$$EU(Gamble\,A) = (0.6 \times £2) + (0.4 \times £4) = 2.8$$
$$EU(Gamble\,B) = (0.6 \times £14) + (0.4 \times £0) = 2.8$$

In a series of experiments, Edwards found that most people have definite preferences between bets of equal expected value. Compared to a good chance of winning a small amount they preferred a long shot of winning a large amount – provided there was no chance of losing very much. People strongly avoided gambles that involved even a low probability of losing a lot. Edwards concluded that SEU was not a guide for

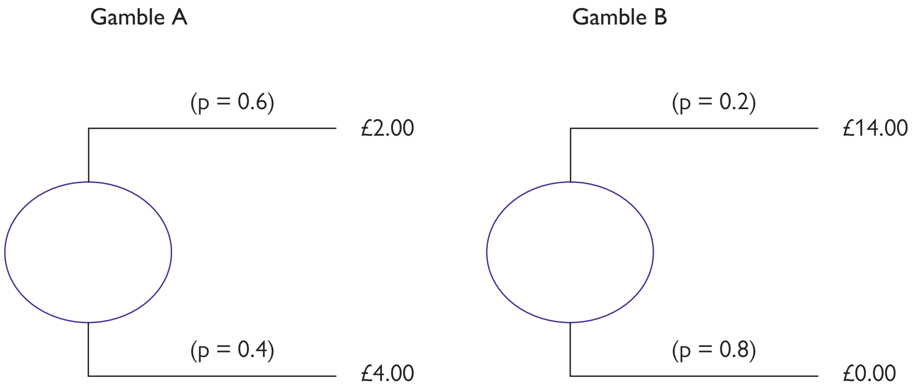


FIGURE 11.2 Two simple gambles of equal expected value.

their choices between gambles. Later, Lichtenstein *et al.* (1969) found that expected value is irrelevant even when the concept was carefully explained to respondents.

3.1 The 'preference reversal phenomenon'

Far worse was to come for SEU, however, when the preference reversal phenomenon was discovered. Slovic and Lichtenstein (1968) had noticed that ratings of a gamble's attractiveness, as well as choices between pairs of gambles, were strongly influenced by the *probability* of winning and losing. Yet when asked how much they would be willing to pay in order to take the gamble, or the smallest amount they would be willing to sell the gamble for, people were more heavily influenced by the *amounts* that could be won or lost. Lichtenstein and Slovic (1973) realized that if there was a different basis for *choosing* than for *valuing*, it should be possible to construct pairs of gambles so that people would prefer A to B but pay more for B than A. They were able to demonstrate this effect in a series of studies – including one conducted with real gambles in a Las Vegas casino. Typically, one bet would have a high probability of winning a modest amount (called the 'p bet') while the other would offer a lower probability of winning a higher amount (called the '$ bet'):

p bet: 11/12 chance of winning 12 chips

1/12 chance of winning 24 chips

$ bet: 2/12 chance of winning 79 chips

10/12 chance of losing 5 chips

These two gambles were chosen equally often by the casino subjects; however, the $ bet received a higher selling price about 88 per cent of the time. Among those choosing the *p* bet, 87 per cent gave a higher selling price to the $ bet. So, people value the $ bet more highly than the *p* bet, but don't prefer the $ bet any more than the *p* bet. From a rational perspective this is a hopeless pattern of behaviour.

The finding, replicated numerous times since (Slovic, 1995), clearly poses a major threat to the SEU view of rational choice. Two economists, Grether and Plott (1979, p.623), realizing that: 'it suggests that no optimisation principle of any sort lies behind even the simplest of human choices', conducted a series of studies 'designed to discredit the psychologists' works as applied to economics'. However, even after controlling for all the economic explanations of the phenomenon that they could find – including that the experiment be conducted by economists rather than psychologists ('Psychologists have the reputation for deceiving subjects', p.629) – the reversals persisted.

3.2 Causes of anomalies in choice

Why do preference reversals occur? Slovic (1995) summarized the evidence in favour of a 'scale compatibility hypothesis'. The idea is that the weight of an option attribute in judgement or choice is influenced by its compatibility with the response mode. As economic value is expressed in terms of money, subjects find it easier to use the monetary aspect of the gamble to set the *value* of the gamble. However, when asked which gamble they *prefer* subjects have no similarly compelling reason to weight the monetary aspect of the gamble to determine their choice. You should note that this explanation does not depend on the presence of risk or uncertainty, and indeed Tversky *et al.* (1990) have demonstrated preference reversals for options where no risk is present.

3.2.1 The 'prominence effect'

To account for another aspect of preference reversals, Tversky *et al.* (1988) identified a specific instance of the compatibility effect, which they termed the **prominence effect**. Slovic (1975) had observed that, after earlier adjusting the payoffs of two gambles so as to make them equally valuable, people did not randomly choose between them but typically chose the gamble with the higher likelihood of winning. Tversky *et al.* (1988) suggested that the more prominent (or important) attribute would weigh more heavily in making a choice than in a matching task (as explained below). For example, in one problem, respondents were asked to imagine two programmes being considered by a transport ministry for dealing with traffic accidents in a country where 600 people are killed every year. Both programmes were described in terms of their annual costs and the expected annual number of casualties that would result if each was introduced. For the choice task, people were asked to choose between the following two options:

Choice task

Option A	570 casualties	Cost $12 million
Option B	500 casualties	Cost $55 million

Of those who took part in the experiment, 67 per cent preferred B to A – note that this implies that the difference in casualties (70) is more important than the difference in costs ($43 million).

Other respondents did a matching task, where they had to fill in the missing value so as to make the two programmes equally attractive (as below):

Matching task

Option A	570 casualties	Cost $12 million
Option B	500 casualties	Cost $?

The typical matching value was less than $55 million – indeed, only 4 per cent of respondents gave a value higher than $55 million.

Plainly, the trade-off between attributes is different with matching than with choice. Why? Tversky *et al.* argued that *choice* invites more *qualitative* reasoning – people select the option that is superior on the most important attribute (lives saved). This is cognitively simpler, easier to justify, and resolves the conflict between the two attributes – albeit by effectively ignoring it. Matching, however, entails a more quantitative assessment. The matching task cannot be performed at all without paying attention to the values of both attributes and their relative importance.

Real world choices often resemble either matching or choice tasks. For example, you might ask yourself what is the most you are prepared to pay when shopping for a particular item (as in the matching task), or you could ask yourself whether you are willing to pay the advertised price for the item (as in the choice task). The evidence suggests that the decisions will tend to diverge. Similar effects may well affect budget setting and resource-allocation decisions. Comparing budget allocation (matching) with budget cutting (choice), the prominence hypothesis suggests that, when forced to choose what items to cut from a hospital budget, health provision (the most important attribute) may fare better than (say) staff pay.

3.2.2 Choosing and rejecting options

Shafir (1993) has shown that choosing one of two items is not the complement of rejecting one of the two items. Sometimes when deciding between two options, people both select and reject the same option. When we are trying to select an option we tend to focus on positive features and when we are looking for reasons to reject an item we tend to focus on negatives. Thus, items that have obvious positive features will be selected over items that do not. Similarly, items that have obvious negative features will be rejected before items that do not. It seems that rather than ranking options in order, as mandated by SEU, people look for *reasons* for their decisions. This has led to the proposal of a reason-based theory of choice (Shafir *et al.*, 1993), according to which reasons for choosing are more influential when we choose rather than reject, and reasons for rejecting are more influential when we reject rather than choose.

Failure to resolve conflict in choices can also be revealing of reasoning. The economist, Thomas Schelling, tells of an occasion when he went to buy an encyclopaedia for his children (see Shafir *et al.*, 1993). At the bookshop he was presented with two attractive encyclopaedias and, finding it difficult to choose between them, went home with neither – despite feeling that he would have happily bought either if it had been the only one available. Unresolved conflict can cause people to defer choosing because they lack a clear reason to select either option.

3.2.3 The 'evaluability principle'

Difficulty in interpreting value is addressed in a study conducted by Hsee (1998), who has developed the notion of evaluability to explain a type of preference reversal that occurs when items are evaluated separately or jointly. For example, if shopping for a piano in a musical instrument shop you might compare several pianos. At an auction or second-hand shop, however, you might have to consider a single piano. Hsee argues that attributes vary in how easy or difficult they are to evaluate, and that their evaluability varies according to whether options are considered in isolation or in relation to other options. In one experiment Hsee asked people to assume they were music students looking for a used music dictionary. In the joint-evaluation condition, participants were shown two dictionaries, A and B (see Table 11.3) and asked how much they would be willing to pay for each. Willingness-to-pay was higher for Dictionary B ($27) than A ($19), presumably because of its greater number of entries. However, in the single evaluation condition, when one group of participants evaluated only A and another group evaluated only B, the mean willingness-to-pay was higher for A ($24) than B ($20).

TABLE 11.3 Attributes of two dictionaries in Hsee's (1998) study

	Year of publication	Number of entries	Any defects?
Dictionary A	1993	10,000	No, it's like new
Dictionary B	1993	20,000	Yes, the cover is torn; otherwise it's like new

Hsee explains this reversal by means of the evaluability principle. He argues that, without a direct comparison, the 'number of entries' attribute is hard to evaluate because the evaluator does not have a precise notion of *how good* or *how bad* 10,000 (or 20,000) entries is. However, the 'defects' attribute is evaluable because it translates easily into a precise good/bad response – most people find a defective dictionary unattractive and a like-new one attractive – and thus it carries more weight in the independent evaluation. Under joint evaluation, however, the buyer can see that B is far superior on the more important attribute, number of entries. Thus, the 'number of entries' attribute becomes *evaluable* through the comparison process.

SUMMARY OF SECTION 3

- Preference reversals illustrate that SEU fails to predict aspects of human decision making.
- The prominence and evaluability of attributes, and whether a task involves choosing or rejecting, have been shown to influence people's choices, effects not predicted by SEU.

4 PROSPECT THEORY

One important general conclusion that follows from these demonstrations of anomalies in choice is that people don't have a set of pre-existing stable values (that is, preferences) that they simply apply to choice situations. What is evident is that decisions change because the underlying bases of decisions change according to the demands of the decision task and the nature and context of the information presented. The unstable nature of preferences raises difficult – perhaps even unsolvable – questions regarding people's preferences. If different procedures for eliciting preferences elicit different choices, then how can preferences be defined and how should they be measured?

Kahneman and Tversky (1979) have proposed a descriptive model for decision making under risk, called prospect theory, which explains many of the phenomena that cannot be accounted for by SEU. Unlike SEU, prospect theory does not define ideal choices. It is a descriptive not a normative theory, intended to account for human choices. Prospect theory is essentially an adapted version of SEU, which is modified so as to account for the observed discrepancies with SEU. Prospect theory identifies two phases to the choice process:

1. In the editing phase, the decision problem is represented; 'negligible' components may be discarded and a reference point is used to enable decision outcomes to be construed as 'gains' or 'losses'.
2. In the second evaluation phase, attitudes towards risks involving gains and losses are used to evaluate the identified prospects.

Prospect theory proposes that people evaluate decision outcomes in terms of gains or losses from a neutral reference point. Figure 11.3 shows how people are thought to value gains and losses. The horizontal axis to the right of the origin shows objective gains ($); as they increase the subjective value of the gains (v($)) also increases but with a diminishing slope. This illustrates the fact that, for example, the psychological difference between $0 and $10 is greater than that between $100 and $110. Notice that there is a similar effect for losses – the slope to the left of the origin shows that losses also diminish in a similar fashion. As the slope is not uniform, your *attitude* to risks varies as a function of where you see yourself on the curve. For contemplating gains (to the right of the origin) decisions will tend to be *risk averse* – most people decline to risk a gain of $10 for a 50 per cent chance of winning $20. By contrast, with losses (to the left of the origin) decisions tend to be *risk seeking* – in order to avoid a sure loss of $10, most people would be tempted to risk a 50 per cent chance of losing $20.

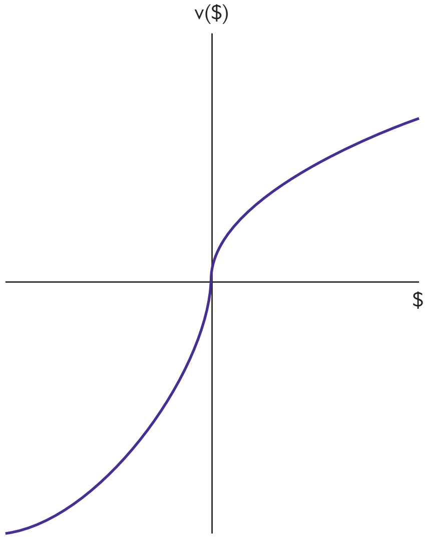


FIGURE 11.3 The value function in prospect theory showing the subjective value of a gain or loss as a function of the actual or objective amount of gain or loss.

4.1 Prospect theory and 'loss aversion'

Another critical feature of the value function is that the curve is steeper for losses than for gains. This models the observation of loss aversion – that people feel losses more than they do gains of equivalent value.

Famously, the economist Paul Samuelson once offered a bet to an economist colleague. They would flip a coin and if the colleague won he would get $200; if he lost he would have to pay Samuelson $100. The colleague, claiming he would feel the $100 loss more than the $200 gain, turned the bet down but mentioned that if Samuelson would play the bet 100 times he would play. (You might have noticed that this pair of preferences is paradoxical with respect to SEU – anyone declining one gamble should not accept any number of plays of the same gamble; Samuelson, 1963.) One consequence of loss aversion is the 'endowment effect', whereby mere possession of an item leads to higher valuation of it than if it were not in one's possession – which may be why so many people accumulate so many possessions that they don't need but can't bear to part with. Knetsch and Sinden (1984) endowed subjects with either a lottery ticket or with $2. Then each subject was offered an opportunity to trade – the lottery ticket for the money, or vice versa. Few subjects chose to switch: 30 per cent of those given $2 bought the ticket, while only 14 per cent of ticket-holders sold their ticket for $2. Kahneman *et al.* (1990) found that randomly assigned owners of a coffee mug required significantly more money to part with their possession (around $7) than randomly assigned buyers were willing to pay to acquire it (around $3); owners' loss of the mug loomed larger than buyers' gain of the mug. Those given lottery tickets or mugs seemed to like them better than those who were given money. Loss aversion also explains resistance to change – a general reluctance to depart from the status quo – because the disadvantages of departing from it loom larger than the advantages of the alternatives (Samuelson and Zeckhauser, 1988).

4.2 Prospect theory and probability weighting

Another 'attitude' applied in the evaluation phase is that probability is distorted. Probabilities (p) are replaced by decision weights (π(p)). Note that this distortion does not apply to the *judgement* or *estimation* of probability – prospect theory posits that people choose as if they give low-probability events more weight than they deserve according to their objective probabilities of occurrence, and as if they give high-probability events less weight than they deserve. Figure 11.4 shows that low probabilities (except zero, which is given zero weight) are over-weighted; note that the lower end of the curve is above the diagonal dotted line. Moderate and high probabilities (except certainty, which is given the correct weight of 1) are under-weighted; note that the upper end of the curve is below the diagonal dotted line.

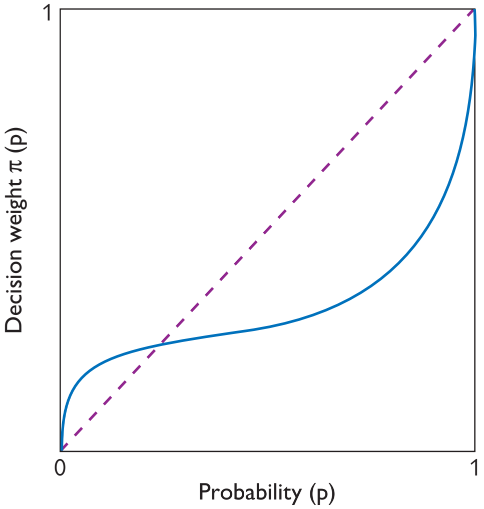


FIGURE 11.4 The weighting function in prospect theory showing the weighted probability π(p) given to probabilities p varying between 0 and 1.

The weighting function of prospect theory accounts for the behaviour observed in the Allais paradox. Because people weight probabilities that are just below certainty far less than they should, they correspondingly give certainty a *relatively* very high value: 100 per cent certainty is weighted a great deal more than 99 per cent. Moreover, very small probabilities are over-weighted – referring back to Allais, we can see that people would worry disproportionately about the 1 per cent chance of not winning. This feature of prospect theory – that small probabilities receive 'too much' weight – explains an otherwise puzzling pattern of risky preferences: many people buy lottery tickets with tiny chances of winning *and* buy insurance *against* tiny chances of loss.

4.3 'Framing' effects

Kahneman and Tversky have reported numerous experiments demonstrating phenomena not anticipated by SEU but predicted by prospect theory. For example, the idea that gains and losses are determined by application of a reference point predicts **framing effects**. Tversky and Kahneman (1981) asked respondents to imagine that the USA was preparing for the outbreak of an unusual disease expected to kill 600 people. Two alternative programmes had been proposed to combat the disease:

- If Programme A is adopted, 200 people will be saved.
- If Programme B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds chance that no people will be saved.

You should note that the options are described in terms of gains – the number of lives that might be saved. Of the respondents, 72 per cent chose Programme A and 28 per cent chose Programme B – definitely saving 200 lives is seen as more attractive than a one-third chance of saving 600 lives. For gains, as we saw in Figure 11.3, people are risk averse – as a result, gains that are certain are more attractive than a gamble of equal expected value.

A second group of respondents was presented with a different description of the two programmes:

- If Programme C is adopted, 400 people will die.
- If Programme D is adopted, there is a one-third probability that nobody will die and a two-thirds chance that 600 people will die.

In this case, only 22 per cent of respondents chose Programme C while 78 per cent chose Programme D. Of course, Programmes C and D are identical to Programmes A and B except that now the outcomes are 'framed' in terms of the numbers of lives that might be lost. Framed as a loss, the same risky option becomes more popular than the riskless option (a clear violation of the invariance axiom that you met earlier). The reversal of preference can be explained by the change of the reference point in conjunction with the shape of the value function. With gains, the reference point is defined by what will happen if nothing is done: 600 dead. Programme A looks attractive as it definitely saves 200 while Programme B risks a two-thirds chance of saving nobody. The relative over-weighting of certainty will also contribute to the relative attractiveness of the sure gain of Programme A. In the domain of loss, the reference point is defined by the present: nobody has yet died. Programme D is more attractive as, because of the shape of the value function 600 deaths are not seen as substantially worse than 400, and it offers a chance that nobody will die.

Prospect theory is supported by a large number of laboratory studies as well as data from field studies (see Camerer, 2000), which avoid some of the doubts that economists have about laboratory studies (recall Grether and Plott's 1979 scepticism about preference reversals). For example, Camerer *et al.* (1997) studied New York cab drivers. Many said that they set an earnings target for the day and quit once they met it. This sounds sensible: target setting can be motivating for unpleasant or tedious activities, like exercise or exam revision. However, as a result, the drivers quit earlier on days when it was easier to get fares (more people want cabs on rainy days, for example) – which is quite counter to their economic interests. If instead they had quit early on slow days and worked longer hours on good days, they could have earned 8 per cent more. Camerer *et al.* attribute this to loss aversion: drivers hate to quit before they reach the target (a 'loss') but, once they reach it, they aren't as enthusiastic about trying to go beyond, (a 'gain'). The cab drivers' irrationality doesn't just punish themselves: on busier days it will be even harder for customers to find a cab.

4.4 Prospect theory and experience-based decisions

Imagine being asked to choose between the following options: (A) A 90 per cent chance of winning $0 and a 10 per cent chance of winning $10, or (B) $1 for sure. What would you choose? Although both options have equal expected value ($1) most people prefer to take the gamble – a result that prospect theory explains by assuming that people over-weight small probabilities. However, recently evidence has emerged that if people learn about the likelihoods of decision outcomes via experience – instead of by reading descriptions – they have different preferences. In experiments where subjects learned about each option by pressing unmarked buttons and sampling the outcomes until they felt ready to choose, people preferred the $1 for sure (e.g. Hertwig *et al.*, 2004). This looks like under-weighting of small probabilities – the opposite of prospect theory's prediction.

The reason for this difference is the subject of much debate – see Rakow and Newell's (2010) review – with some calling for a different theory of choice for experience-based decisions (Hertwig *et al.*, 2004) and others arguing that the difference is due to insufficient sampling of rare outcomes in experience-based decisions (Hadar and Fox, 2009) or memory limitations (see Rakow *et al.*, 2008). Others claim that sampling and memory effects cannot account for it (Ungemach *et al.*, 2009). Is this a problem for prospect theory? The answer is not yet clear but it is obviously important; although most experiments have tested described decisions, most of our choices in life are made without the benefit of described probabilities and we must rely on experience.

SUMMARY OF SECTION 4

- Prospect theory can account for an enormous range of observed anomalies in choice, both in laboratory experiments and in field data representing real-life decisions.
- People feel losses more than they do gains of equivalent value.
- Framing of options influences preference patterns.
- Experience-based decisions show under-weighting of small likelihoods, counter to the predictions of prospect theory and yet to be satisfactorily accounted for.

5 JUDGEMENT UNDER UNCERTAINTY

As noted earlier, judgement and choice can be distinguished on the basis that judgements are what underlie choices. To compute the ideal choice, the normative theory requires inputs of value and likelihood, which typically will be judgements by decision makers. Plainly, evaluating likelihood is a crucial prerequisite for effective decision making under uncertainty. For this reason, research on judgement of likelihood has a particular significance.

5.1 Judging probabilities and Bayes' Theorem

In the 1960s, Ward Edwards and his colleagues conducted a number of studies using what were called the **book bag** and **poker chip paradigms**. A typical experiment would involve two opaque bags. Each bag contained 100 coloured poker chips in different but stated proportions of red to blue. Suppose Bag A contains 70 red and 30 blue chips, while Bag B contains 30 red and 70 blue chips. The experimenter first chooses one bag at random and then draws a series of chips from it. After each draw, the poker chip is replaced and the bag well shaken before the next chip is drawn. The subjects' task is to say how confident they are – in probability terms – that the chosen bag is A, containing predominantly red chips, or B, containing predominantly blue chips.

Bayes' Theorem can be used to calculate how the objective probabilities change after each piece of new information, and so can be used to evaluate human performance (see Box 11.2). Where the two competing

BOX 11.2 RESEARCH STUDY Using Bayes' Theorem

So, how good are people at judging probabilities? In order to work this out, we can try using an objective standard to which we can compare human performance. One early benchmark used for this purpose was Bayes' Theorem (which you will also meet in Chapter 12), a mathematical formula used for combining probabilities; it is of interest to decision makers as it can be used as a normative theory for how degrees of belief in a hypothesis might be revised in the light of new information.

Bayes' Theorem states that the odds of a hypothesis being correct in the light of new information (**posterior odds**) is the product of two elements: (1) the **prior odds** (the initial odds) of the hypothesis being correct *before* the information is observed, multiplied by (2) the **likelihood ratio** – the ratio of the probabilities that, given the information, the hypothesis is correct (*H*) or incorrect ($\bar{H}$).

$$\frac{p(H|D)}{p(\bar{H}|D)} = \frac{p(D|H)}{p(D|\bar{H})} \times \frac{p(H)}{p(\bar{H})}$$

(H = hypothesis; D = datum)

You should note that $p(H|D)$ stands for the probability that the hypothesis is true, given that the datum or information *D* is true (read the vertical line as 'given'). Then, reading the formula from left to right, the three terms are:

1 The posterior odds ratio of *H* and $\bar{H}$ (not *H*) being true given *D*.
2 The likelihood ratio, representing the information value of *D* (the datum or information).
3 The prior odds of *H* and $\bar{H}$ (not *H*) being true before *D* is known.

If the probability of observing *D* when *H* is true is different from the probability of observing *D* when $\bar{H}$ is not true, then the information is diagnostic and the posterior odds will be different to the prior odds. The equation can be applied to any pair of competing hypotheses (A and B) by replacing *H* and $\bar{H}$ with A and B.

hypotheses are that the bag is A and the bag is B, and the information is the drawing of a red chip, Bayes' Theorem gives the following equality:

$$\frac{P(A|RED)}{P(B|RED)} = \frac{P(RED|A)}{P(RED|B)} \times \frac{p(A)}{p(B)}$$

where $p(\text{A}|\text{RED})$ stands for the probability that the bag is A, given that a red chip has been drawn.

Prior to drawing any chips, the probability that the bag is A and that the bag is B is 0.5 for both; if we draw one red chip the likelihood of this is 0.7 for Bag A and 0.3 for Bag B. We can substitute these values into Bayes' Theorem:

$$\frac{p(A|RED)}{p(B|RED)} = \frac{0.7}{0.3} \times \frac{0.5}{0.5}$$

$$\frac{p(A|RED)}{p(B|RED)} = \frac{0.35}{0.15}$$

Therefore, the odds for Bag A over Bag B after drawing one red chip are 0.35:0.15 or, converting these into percentage probabilities, 70 per cent for Bag A and 30 per cent for Bag B. Bayes' Theorem can then be used after each subsequent drawing of a chip to calculate how these probabilities should change.

Thus, for example, following the drawing of the red chip and prior to drawing a second chip, our new prior odds ratio is $\frac{0.7}{0.3}$. If we replaced the red chip and shook the bag so that the probabilities of drawing a red and blue chip remained as before, the impact of drawing a second chip would now be applied to our new prior ratio to compute another posterior odds ratio. If it was yet another red chip, belief in Bag A would be $\frac{0.7}{0.3} \times \frac{0.7}{0.3} = \frac{0.49}{0.09}$ – roughly 85%:15% – even more favourable for Bag A. If the second chip was blue, the information for each bag would be equivocal – our new posterior odds ratio would be $\frac{0.3}{0.7} \times \frac{0.7}{0.3} = \frac{0.21}{0.21}$ that is, 50:50 odds that it is Bag A or Bag B.

A crucial aspect of the logic of these studies is that the experimenter is able to compare the subjective probabilities estimated by subjects with the objective probabilities calculated using Bayes' Theorem. All of the information required as inputs to Bayes' Theorem is explicit and unambiguous. Ironically, though, this underplays the importance of the *subjectivity* of the probabilities. Because the experimenters assumed that they could objectively compute *the* correct answer, they reasoned that the subjective probabilities should be the same for all subjects faced with the same evidence. Calculating the objective probabilities and using them as a comparison was absolutely necessary in order to be able to assess the accuracy of people's judgements. However, it also indicates the artificiality of this kind of task, and is at the root of the difficulties that were to emerge with interpreting subjects' behaviour.

5.2 Does Bayes' Theorem describe human judgement?

The experiments conducted with the procedure discussed above produced ample evidence that human judgement under these conditions is not well described by Bayes' Theorem. Although subjects' revisions of their probability judgements were proportional to the values calculated from Bayes' Theorem, they did not revise their opinions sufficiently in the light of the evidence – a phenomenon that was labelled **conservatism**. The clear suggestion was that human judgement was poor, although there was some debate as to the precise reason for this. Perhaps it was due to a failure to understand the impact of the evidence or to an inability to combine the probability estimates according to Bayes' Theorem.

Aside from the theoretical interest in these possibilities, there were practical implications of this debate. If people are good at assessing probabilities but poor at combining them (as Edwards, 1968 suggested) then perhaps they could be helped; a relatively simple remedy would be to design a support system that took the human assessments and combined them using Bayes' Theorem. However, if people were poor at assessing the component probabilities, then there wouldn't be much point in devising systems to help them to combine them. 'Garbage in – garbage out' was a popular aphorism for summarizing this possibility.

However, before any firm conclusions were reached as to the cause of conservatism, the research exploring the phenomenon fizzled out. Two reasons for this can be identified. One cause (which we will consider in Sections 5.3 and 5.4) was the emergence of research into heuristics and biases and, in particular, the discovery of what Kahneman and Tversky (1973) called base-rate neglect. Before this development, however, growing disquiet was being voiced about the validity of book bag and poker chip experiments for assessing judgement.

Several studies had shown that quite subtle differences in the way that the tasks were presented to subjects resulted in considerable variability in the amount of conservatism. For example, the **diagnosticity** of the data seemed an important variable. Diagnosticity of information means how much impact it has on opinion revision. Diagnosticity is indicated by the likelihood ratio. Imagine, instead of our two bags with a 70/30 split in the proportions of blue and red poker chips, the bags contained 51 chips of one colour and 49 of the other. Clearly, two consecutive draws of a red chip would not be very diagnostic as to which of the two bags was being sampled. Phillips and Edwards' (1966) experiments showed that the more diagnostic the information, the more conservative was the subject. But when the information was very weakly diagnostic, as in this particular example, human probability revision, far from being conservative, was too extreme.

Another important factor was how the information was presented. Presenting information all at once or one bit at a time is irrelevant according to Bayes' Theorem, but Peterson *et al.* (1965) found that presenting one item of information at a time, eliciting revisions after each item, produced less conservatism than giving the same information all in one go. Pitz *et al.* (1967) described this as an **inertia effect**: if an initial sequence of information favoured one of the hypotheses under evaluation, subjects tended not to reduce their belief when confronted with later conflicting information.

DuCharme and Peterson (1968) investigated probability revision in a situation they considered nearer to real life than the standard paradigm. Most experiments, they complained, usually restricted information to one of two discrete possibilities (red or blue chip). In the real world, information leading to revision of opinion does not have discrete values but varies along a continuum. They gave their subjects the task of deciding which population was being sampled from – males or females – on the basis of the information given by randomly sampling heights from one of the populations. Using this task, DuCharme and Peterson found conservatism reduced to half the level found in the more artificial tasks. They concluded that this was due to their subjects' greater familiarity with the height distributions underlying their task.

Winkler and Murphy (1973) expressed further doubt concerning the validity of the conclusions from the book bag and poker chip paradigms. They argued that the standard task differed in several crucial aspects from the real world:

1. The pieces of evidence usually presented to subjects are conditionally independent. That is, knowing one piece of information does not change the likelihood of the other: producing one red chip from a bag, and then replacing it, does not affect the likelihood of drawing another red chip. However, in real-world situations this assumption often does not make sense. For example, someone trying to discriminate hostile from friendly aircraft might spot an aircraft flying a non-standard route that fails to respond to radio signals. Flying off course and failing to respond are not independent – both could be caused by equipment failure. So, after observing one we should be less influenced by the other.

   Winkler and Murphy argued that in many real-world situations, lack of conditional independence of the information renders much of it redundant. In the standard tasks, subjects may have treated the information as if it was conditionally dependent and so one possible explanation for conservatism is that subjects are behaving much as they would do in familiar situations that involve redundant information sources.

2. In most experiments, the contents of the bags are fixed but in reality our hypotheses are not always constant; indeed, evidence may cause us to change the set of hypotheses under consideration.
3. In reality, information may be somewhat unreliable and therefore less diagnostic than the perfectly reliable colours of the poker chips.
4. Typical experiments offer very diagnostic evidence – clearly favouring one hypothesis – whereas in reality evidence is very often weakly diagnostic. Again, the result of generalizing from experience may be the appearance of conservatism. You will recall Phillips and Edwards' (1966) discovery that probability revision was too extreme with very weakly diagnostic evidence.

Winkler and Murphy (1973) concluded that 'conservatism may be an artifact caused by dissimilarities between the laboratory and the real world'.

5.3 Heuristics and biases

From the early 1970s Kahneman and Tversky (e.g. Kahneman and Tversky, 1972; Tversky and Kahneman, 1974) provided a plethora of demonstrations of human judgemental error and linked these to the operation of a set of **mental heuristics** – mental rules of thumb – that they proposed the mind uses to simplify the process of judgement. These foibles, they argued, indicated that the underlying processes of judgement were not normative (e.g. did not compute probabilities using Bayes' Theorem) but instead used simpler rules that were easier for the brain to implement quickly.

The logic of their empirical research was to infer the characteristics of the mental processes underlying judgement by studying persistent biases – those not due to inattention or fatigue. The idea, spelled out in Kahneman *et al.* (1982), is that, due to limited mental processing capacity, strategies of simplification are required to reduce the complexity of judgement tasks and render them tractable by the kind of mind that people have. Accordingly, the principal reason for interest in judgemental biases was not merely that subjects made errors but that the errors revealed how people made use of relatively simple but error-prone heuristics for making judgements.

5.3.1 The 'representativeness' heuristic

The **representativeness** heuristic is used to determine how likely it is that an event is a member of a category by considering how similar or typical the event is to the category (remember the similarity-based approach to categorization discussed in Chapter 5?). For example, people may judge the likelihood that a given individual is a librarian by the extent to which the individual resembles a 'typical' librarian. This may seem a reasonable strategy but it neglects consideration of the relative prevalence of librarians in society as a whole: the so-called **base rate**. We have seen that Bayes' Theorem prescribes that prior likelihood is an important component when assessing the impact of new information. So, when given information about an individual, the chances that he or she is a member of a profession will still be influenced by the prior likelihood – or base rate – for that profession. Knowing that someone regularly works in the British Library might increase your belief that they are a famous writer, but it is still more likely that she or he is a librarian because there are more librarians than famous writers. Tversky and Kahneman found that when base rates of different categories varied, judgements of the occupations of described people were correspondingly biased – due to base-rate neglect. People using the representativeness heuristic for forecasting were employing a form of stereotyping in which similarity dominated other cues as a basis for judgement.

In Kahneman and Tversky's (1973) experiments demonstrating neglect of base rates, subjects were found to ignore information concerning the prior probabilities of the hypotheses – the polar opposite of conservatism. For example, in one study subjects were presented with this brief personal description of an individual called Jack.

> *Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing and mathematical puzzles.*

Half the subjects were told that the description had been drawn from a sample of 70 engineers and 30 lawyers while the other half were told that the description was drawn from a sample of 30 engineers and 70 lawyers. So, the base rate or prevalence of engineers for the two groups was 70 per cent and 30 per cent respectively. However, when asked to estimate the probability that Jack was an engineer, the mean estimates of the two groups were only very slightly different (50 per cent vs 55 per cent). On the basis of such results, Kahneman and Tversky concluded that prior probabilities are largely ignored when individuating information is made available.

Kahneman and Tversky then gave a description designed to be totally uninformative about the profession of the individual:

> *Dick is a 30-year-old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues.*

When contemplating this description, subjects given markedly different base rates produced identical median estimates of 50 per cent. Kahneman and Tversky concluded that base rates were properly utilized when no specific information was given, but that base rates were neglected when even worthless information was provided (as in this example).

Tversky and Kahneman (1983) also invoked judgement by representativeness to explain the **conjunction fallacy**, whereby a conjunction of two events is judged to be more likely than one of those events alone. The fallacy violates a simple principle of probability: the probability of a conjunction A and B can never exceed either the probability of A or the probability of B. Nevertheless, subjects who read a description of a woman called Linda who had a history of interest in liberal causes thought it more likely that she was a feminist bank clerk (i.e. a conjunction – Linda is a feminist *and* a bank clerk) than just a bank clerk, thereby violating the conjunction rule. Of course, though all feminist bank clerks are bank clerks, feminist bank clerks are more *representative* of people interested in liberal causes than bank clerks in general. So, while valid probabilities respect the conjunction rule, judgements of representativeness may not.

5.3.2 The 'availability' heuristic

The **availability** heuristic is invoked when people estimate likelihood or relative frequency by the ease with which instances can be brought to mind. Instances of frequent events are typically easier to recall than instances of less frequent events, so availability will often be a valid cue for estimates of likelihood. However, availability is affected by factors other than likelihood. For example, recent events and emotionally salient events are easier to recollect. It is a common experience that the perceived riskiness of air travel rises in the immediate wake of an air disaster. Judgements made on the basis of availability, then, are vulnerable to bias whenever availability and likelihood are uncorrelated.

5.3.3 The 'anchor and adjust' heuristic

The **anchor and adjust** heuristic is used when people make estimates by starting from an initial value and then adjust it to arrive at their final estimate. The claim is that adjustment is typically insufficient. For instance, one experimental task required subjects to estimate various quantities stated in percentages (e.g. the percentage of African countries in the UN). Subjects communicated their answers by using a spinner wheel showing numbers between 0 and 100. For each question, the wheel was spun and then subjects were first asked whether the true answer was above or below this arbitrary value. They then gave their estimate of the actual value. Perversely, people's estimates were found to correlate with the initial (entirely random) starting point (see Wilson *et al.*, 1996).

5.4 Evaluating the heuristics and biases account

The heuristics and biases research provided a methodology, a vivid explanatory framework, and a strong suggestion that judgement is not as good as it might be. Kahneman and Tversky (1982) made clear that the main goal of their research was to understand the processes that produce both valid and invalid judgements. However, it soon became apparent that 'although errors of judgement are but a method by which some cognitive processes are studied, the method has become a significant part of the message' (Kahneman and Tversky, 1982, p.494). So how should we regard human judgement?

There has been an enormous amount of discussion of Tversky and Kahneman's findings and claims. Researchers in the heuristics and biases tradition have generated shock and astonishment that people seem so bad at judging probability despite the fact that we all live in an uncertain world. Not surprisingly, these claims have been challenged. Some question whether the demonstrations of biases in judgement apply to experts operating in their domain of expertise or merely to student samples. Another argument is that the experimental tasks set to subjects provide a misleading perspective of their competence. A third argument is that the standards for the assessment of judgement are inappropriate.

Consideration of a prominent critique of Tversky and Kahneman's argument is given below.

5.4.1 Representativeness and base-rate neglect

Following Tversky and Kahneman's original demonstration of base-rate neglect, research established that base rates might be attended to more (though usually not sufficiently) if they were perceived as relevant (Bar-Hillel, 1980), had a causal role (Kahneman and Tversky, 1982), or were 'vivid' rather than 'pallid' (Nisbett and Ross, 1980). However, Gigerenzer *et al.* (1988) argued that the variations in base-rate neglect have nothing to do with any of these factors *per se*, but arise because different problems may to varying degrees encourage the subject to represent the problem as a Bayesian revision problem. Just because the experimenter assumes that he or she has defined a probability problem does not imply that the subject will see it in the same way. In particular, subjects may have reasons not to take the base rate asserted by the experimenter as their subjective prior probability. In

Kahneman and Tversky's original experiments the descriptions were not actually randomly sampled (as the subjects were told) but especially selected to be 'representative' of the professions. If subjects suspected that this was the case then they would be entitled to ignore the offered base rate.

Gigerenzer *et al.* (1988) let their subjects experience the sampling themselves. Their subjects examined ten pieces of paper each marked lawyer or engineer in proportion to the base rates. Subjects then drew one of the pieces of paper from an urn and unfolded it so they could read a description of an individual without being able to see the mark defining it as being of a lawyer or engineer. In these circumstances, subjects used the base rates in a proper fashion – base-rate neglect 'disappeared'. However, in a replication where base rates were asserted rather than sampled, Kahneman and Tversky's base-rate neglect was replicated.

Kahneman and Tversky (1996) have argued that a fair summary of the research would be that explicitly presented base rates are generally under-weighted but not ignored. They also pointed out that, in Gigerenzer *et al.*'s (1988) experiment, subjects who sampled the information themselves still produced judgements that deviated from the Bayesian solution in the direction predicted by representativeness. Evidently, then, representativeness is useful for predicting judgements. However, to the extent that base rates are not entirely ignored (Koehler, 1995), the heuristic rationale for representativeness is limited. You will recall that the original explanation for base-rate neglect was the operation of a simple heuristic that reduced the need for integration of information. If judgements in these experiments reflect use of base rates – albeit to a limited extent – it is hard to account for findings by the operation of a simplifying representativeness heuristic.

5.4.2 Frequency and the conjunction fallacy

Tversky and Kahneman (1983) reported evidence that violations of the conjunction rule largely disappeared when subjects were requested to assess the relative **frequency** of events rather than the probability of a single event. Thus, instead of being asked about the likelihood for a particular individual, subjects were requested to assess how many people in a survey of 100 adult males had had heart attacks and then were asked to assess the number who were both over 55 years old *and* had had heart attacks. Only 25 per cent of subjects violated the conjunction rule by giving higher values to the latter than to the former. When asked about likelihoods for single events, however, it is typically the vast majority of subjects who violate the rule. This difference in performance between frequency and single-event versions of the conjunction problem has been replicated several times since (see Gigerenzer, 1994).

Gigerenzer (1994) has suggested that people are naturally adapted to reasoning with information in the form of frequencies and that because of this the conjunction fallacy 'disappears' if reasoning is based on frequencies. This suggests that the difficulties that people experience in solving probability problems can be reduced if the problems require subjects to assess relative frequency for a class of events rather than the probability of a single event. Thus, it is possible that if judgements were elicited with frequency formats then there would be no biases. Kahneman and Tversky (1996) disagree and argue that the frequency format serves to provide subjects with a powerful cue to the relation of inclusion between sets that are explicitly compared or evaluated in immediate succession. When the structure of the conjunction is made more apparent, then subjects who appreciate the constraint supplied by the rule will be less likely to violate it. According to their account, salient cues to set inclusion – not the frequency information *per se* – prompted subjects to adjust their judgement.

To test this explanation, Kahneman and Tversky (1996) reported a new variation of the conjunction problem experiment where subjects made judgements of frequencies but the cues to set inclusion were removed. They presented subjects with the description of Linda and then asked their subjects to suppose that there were 1,000 women who fit the description. They then asked one group of subjects to estimate how many of them would be bank tellers; a second, independent group of subjects to estimate how many were bank tellers and active feminists; and a third group to make evaluations for both categories. As predicted, those subjects who evaluated both categories mostly conformed to the conjunction rule. However, in a between-groups comparison of the other two groups, the estimates for 'bank tellers and active feminists' were found to be significantly higher than the estimates for bank tellers. Kahneman and Tversky argue that these results show that subjects use the representativeness heuristic to generate their judgements and then edit their responses to respect class inclusion where they detect cues to that relation. Thus, they concluded that the key variable controlling adherence to the conjunction rule is not the relative frequency format *per se*, but the opportunity to detect the relation of class inclusion.

Other authors have investigated the impact of frequency information (Evans *et al.*, 2000; Girotto and Gonzales, 2002) and concluded that it is not the frequency information *per se* but the perceived relations between the entities that is affected by different versions of the problem, though this is rejected by Hoffrage *et al.* (2002). We need to understand more of the reasons underlying the limiting conditions of cognitive biases – how it is that seemingly inconsequential changes in the format of information can so radically alter the quality of judgement. Biases that can apparently be cured so simply cannot plausibly be held to reveal fundamental and immutable characteristics of judgement processes. We shall now consider the history of one well-known cognitive bias: overconfidence.

5.5 Overconfidence

In the 1970s and 1980s a considerable amount of evidence was marshalled for the view that people suffer from an overconfidence bias, often assumed to result from information-search strategies and motivation (see Kunda, 1990). Associative memory and confirmatory search strategies could facilitate retrieval of information that is consistent with initial impressions, and those impressions also then bias the interpretation of subsequent ambiguous evidence. Motivational factors could cause or exacerbate the bias. People like to think that they are intelligent and knowledgeable, and they may have reasons for wanting a particular answer to be true. Typical laboratory studies of calibration ask subjects to answer binary questions such as:

> *'Which is further south?' (a) Rome, or (b) New York*

Subjects are required to indicate the answer that they think is correct and then state how confident they are on a probability scale ranging from 50 per cent to 100 per cent (the minimum is 50 per cent since one of the answers is always correct and 50 per cent is the probability of guessing correctly). To be well calibrated, an assessed probability should correspond with the number of correct judgements over a number of assessments. For example, if you assign a probability of 70 per cent to each of ten predictions then you should get seven of those predictions correct. Typically, however, people tend to give **overconfident** responses – their average confidence is higher than their proportion of correct answers. McClelland and Bolger (1994) and Harvey (1997) give comprehensive reviews of this aspect of probabilistic judgement.

Overconfidence has been recorded in the judgements of experts. For example, Christensen-Szalanski and Bushyhead (1981) explored the validity of the probabilities given by physicians to diagnoses of pneumonia. They found that the probabilities were poorly calibrated and very overconfident; the proportion of patients who turned out to have pneumonia was far less than the probability statements implied. Wagenaar and Keren (1986) found overconfidence in lawyers' predictions of the outcome of court trials in which they represented one side. As they point out, it is inconceivable that the lawyers do not pay attention to the outcomes of trials in which they have participated, so why don't they learn to make well-calibrated judgements?

Could the circumstances in which some experts operate impede the proper monitoring of feedback necessary for the development of well-calibrated judgements? A consideration of the reports of well-calibrated experts supports this notion; they all appear to be cases where some explicit unambiguous quantification of uncertainty is routinely made and the outcome feedback is prompt and unambiguous. Doctors and lawyers don't routinely quantify their uncertainty and may have to wait months to discover the outcomes of their judgements, the truth of which may never be revealed.

The most commonly cited example of well-calibrated judgements is weather forecasters' estimates of the likelihood of rainfall (Murphy and Winkler, 1984) but there are others. Keren (1987) found highly experienced tournament bridge players (but not experienced non-tournament players) made well-calibrated forecasts of the likelihood that their bids would be made, and Phillips (1987) reports well-calibrated forecasts of horse races by bookmakers. In each of these three cases, the judgements made by the experts are precise numerical statements and the outcome feedback is unambiguous and received promptly, and so can be easily compared with the initial forecast. Under these circumstances, experts are unlikely to be insensitive to the experience of being surprised; there is very little scope for neglecting, or denying, any mismatch between forecast and outcome.

Following the ideas of Brunswik (1943, 1955) – that cognition is well adapted to people's natural environments – some judgement researchers have argued that overconfidence is an artefact of artificial experimental tasks and the non-representative sampling of stimulus materials. Gigerenzer *et al.* (1991) and Juslin (1994) claim that overconfidence is observed because the typical general knowledge quizzes used in most experiments contain a disproportionate number of misleading items. For example, most people judge wrongly that Rome is south of New York. These authors found that when knowledge items are randomly sampled, the overconfidence phenomenon disappears. Gigerenzer *et al.* (1991) presented their subjects with randomly selected pairs of German cities. When asked to select the biggest and indicate their confidence, overconfidence was not observed.

Erev *et al.* (1994) spotted another misleading source of evidence of overconfidence. They explained that overconfidence might, to some degree, reflect an underlying random component of judgement. When any two variables are not perfectly correlated – and confidence and accuracy aren't *perfectly* correlated – there will be a regression effect. For example, the heights of fathers and their sons are positively – but not perfectly – correlated. Consequently, a sample of the (adult) sons of extremely tall fathers will, on average, be shorter than their fathers *and at the same time* a sample of the fathers of extremely tall (adult) sons will, on average, be shorter than their sons. Thus, depending on how you sampled – either looking at very tall sons or very tall fathers – you could be led to two opposite conclusions about the direction of a (non-existent) difference between the populations. If, as Erev *et al.* argued, confidence and accuracy are both essentially noisy measures of something else (the strength of the underlying information), then regression effects are inevitable. When Erev *et al.* analysed accuracy as a function of confidence (the usual method for studies of overconfidence) they found overconfidence; but when they used *the same data* to analyse confidence as a function of accuracy they found *under*confidence. They concluded that people's confidence judgements reflect lots of random error, but little or no systematic bias.

They also pointed out that such regression effects could explain the conservatism effect discussed earlier, where judgements are less extreme than values calculated using Bayes' Theorem and where confidence is analysed as a function of the 'objective' probabilities. Indeed, to highlight the paradox that overconfidence studies point to judgements being too extreme while conservatism studies conclude that judgements were less extreme than they should be, they call conservatism underconfidence.

So could it really be that all the evidence for overconfidence is merely an illusion created by inappropriate sampling of test items and regression effects? Budescu *et al.* (1997) attempted to measure and control for the regression effects caused by random variation in judgements by presenting the same items (random pairs of large American cities) on several occasions to their subjects. They found that the vast majority of the individuals in their study (87 per cent) were biased towards overconfidence even after the effects of random error in their judgements had been taken into account. As they also used a representative sample of items, both the artefactual sources of overconfidence should have been eliminated.

However, Juslin *et al.* (2000) report a meta-analysis comparing 35 studies where the items for judgement were randomly sampled from a defined domain, with 95 studies where items were selected non-randomly by experimenters. While overconfidence was evident for selected items, it was close to zero for randomly sampled items, which suggests that overconfidence is not simply a ubiquitous cognitive bias. This analysis suggests that the appearance of overconfidence may be an illusion – not one experienced by experimental subjects, but one inadvertently created and suffered by researchers, and so not a cognitive bias in their respondents.

Nevertheless, in other studies, using what is called interval estimates of confidence, people have been found to be persistently overconfident – even when controlling for the way items are sampled and for regression effects from random error in the judgement process. In these studies participants do not choose between binary answers, but rather provide estimates of a quantity in terms of ranges or boundaries that correspond to a given degree of confidence. For example, if asked 'When was Charles Darwin born?' a response might be 'I'm 90 per cent sure the answer is between 1790 and 1820′. Although interval estimates are less studied than binary choices, they are more like the judgements that people commonly make in many contexts. When people decide when to leave for the airport, how much money to change into foreign currency when going abroad, or how much food to buy at the supermarket, they implicitly make judgements about a plausible interval for the time the journey will take, how much they will need to spend, or the rates of consumption and decay. Yaniv and Foster (1997) found that people signal a rough sense of confidence in an interval estimate by varying the precision with which they express information. 'I think it was during the first half of the 19th century' conveys a different degree of confidence from 'I think it was around 1825'.

Klayman *et al.* (1999) and Juslin *et al.* (1999) found that interval estimates are prone to a great deal of overconfidence – even when using representatively sampled items and controlling for random error. When people were asked to provide intervals such that they were *x* per cent sure that the correct answer was within the interval, the answer was inside their interval much less than *x* per cent of the time. If, for example, people are asked to indicate confidence intervals for their judgements about when various items were invented (e.g. the ball point pen) that should include the true value with probability 0.9, the proportion of intervals including the true value tend to be far below 0.9, often closer to 0.4 or 0.5. An overconfidence score is determined as the probability for the interval minus the hit rate for that interval. For example, if a subject used the 90 per cent confidence interval for a set of questions and gave ranges that included the true value for only 50 per cent of the items, they would have an overconfidence score of 90 – 50 = 40 per cent, or 0.4.

A number of subsequent studies have confirmed what is called 'format dependence' – variation in overconfidence depending on the choice of response format and greater overconfidence when giving confidence intervals (Soll and Klayman, 2004; Winman *et al.*, 2004). Because confidence intervals are often made by experts and the difference in expected return for a probability 0.9 or 0.4 can be enormous, the phenomenon is not only theoretically puzzling but of profound applied relevance. Curiously, however, when participants are provided with the same intervals and asked to assess the probability that the value falls within the interval, the overconfidence bias tends to diminish or even disappear (Winman *et al.*, 2004). Of course the method of measurement should not affect underlying confidence, which in itself suggests that the problem is not inappropriate confidence but an artefact of measurement.

It is worth recording that experts are not immune to overconfidence in confidence interval judgements. McKenzie *et al.* (2008) examined both information technology (IT) experts' and novices' judgements about the IT industry – e.g. one item asked 'How much money did US businesses spend on software last year?' Experts reported intervals that had midpoints closer to the true value – which increased hit rate – and that were narrower (i.e. more informative) – which decreased hit rate. The net effect was that overconfidence remained largely the same.

Significant theory developments can account for many – though not yet all – of the observations by invoking sampling and random error rather than systematic biases in information processing. For example, Juslin *et al.*'s (2007) *naïve sampling model* can account for overconfidence with intuitive confidence intervals, yet relatively accurate judgement with other probability judgement formats. The model assumes that people correctly represent the uncertainty implied by the samples of information they experience, but naïvely assume that sample properties reflect population properties. To see how it works, imagine that the 75 per cent confidence interval for, say, the number of inhabitants of Thailand may be formed by retrieving a sample of Asian countries with a known number of inhabitants and reporting an interval that includes 75 per cent of the values in this sample. Thus if I knew the populations of four Asian countries and that three of them had populations between 20 and 70 million I would use this as my 75 per cent confidence interval. Overconfidence arises because the samples one experiences will tend to underestimate the variance of the distribution of populations – for example, in the extreme case where you repeatedly sample just twice from a distribution on a continuous dimension, *all* samples will underestimate the complete range of the distribution, except one (i.e. the one actually containing the maximum and minimum values of the distribution). The four Asian countries whose populations I knew might well not reflect the variance of the populations of all Asian countries.

Other judgement formats won't encounter this difficulty – thus, to assess your percentage confidence that Thailand has more than 25 million inhabitants you may, for example, retrieve a sample of Asian countries with a known number of inhabitants and assess the proportion of those that exceed 25 million. Assuming a representative sample and no regression effects, the process should produce accurate confidence judgements – exactly what has been reported (Gigerenzer *et al.*, 1991; Juslin, 1994; Juslin *et al.*, 2000; Klayman *et al.*, 1999).

By contrast Soll and Klayman (2004) propose that the overconfidence bias is explained by selective and confirmatory search of memory. They report a number of results that are hard to accommodate in a sampling error framework. For example, men were almost twice as overconfident as women; women's estimates were not more accurate but their intervals were more than 50 per cent wider than men's. As Klayman *et al.* (1999) put it, whether judgements are overconfident depends on how, what, and whom you ask.

SUMMARY OF SECTION 5

- There have been many demonstrations of human judgemental error, linking errors to the operation of various heuristics and biases.
- There is evidence that many of these reported biases disappear when the wording of the problems is changed. Whether this reflects sensitivity to particular cues or adaptivity to frequency information formats is the subject of debate.
- Overconfidence research shows a significant change in theorizing. Early explanations invoked the idea of systematic bias in cognitive processes. More recent theories emphasize that overconfidence may not arise primarily from biased processing *per se*, but from inabilities to correct for the effects of sampling and the effects of random – not systematic – error.

6 FAST AND FRUGAL THEORIES OF DECISION MAKING

Another approach to judgement and choice that has recently emerged tests the efficacy of heuristics on information occurring 'in the wild' rather than on specially contrived laboratory problems. Gigerenzer and Goldstein (1996) comparatively evaluated the performance of a set of different decision strategies. Instead of focusing on violations of normative rules, they produced a measure of the efficacy of simple mental strategies for judgement by measuring the number of correct inferences that different strategies made. The class of simple models that Gigerenzer and Goldstein tested were what they called **fast and frugal** heuristics: 'frugal' because these heuristics used just one piece of information in order to make decisions; 'fast' because they didn't attempt any sort of integration of different bits of information prescribed by such normative procedures as SEU or Bayes' Theorem. By the standards of classical rationality enshrined in normative rules, the mental strategies that Gigerenzer and Goldstein considered look very primitive. Indeed, they were quite explicit about the fact that the simple heuristics that they tested violate basic axioms such as transitivity. Nevertheless, the proof of the pudding is in the eating; as we have seen, people and bees violate transitivity and yet manage to get by.

The inspiration for this exercise was Simon's (1956) idea of **bounded rationality**. Simon emphasized that, due to its limited capacity, human information processing would be obliged to use **satisficing** methods for problem solving (satisficing is an old Northumbrian word meaning to satisfy). Simon used it to describe decision procedures that, while not optimal, reflect the constraints supplied by human information-processing capacity and the opportunities provided by the structure of the environment. Most research on human judgement has focused on the non-optimal nature of simple human information-processing strategies – the importance of the environment structure in determining performance has been overlooked. Nevertheless, we have seen how evidence for both conservatism and overconfidence was undermined by considering how mental strategies might exploit the way information is structured in the natural environment. But could judgement strategies that violate normative rules and utilize just one piece of information possibly be of effective service?

In their study, Gigerenzer and Goldstein used the properties of a set of German cities as information on which to base decisions as to which city was the biggest. Commonly known correlates of city size, such as whether it is the state capital, has a university, has a football club in the top division, or has an inter-city rail station, were the cues that the heuristics could use. One heuristic that Gigerenzer and Goldstein tested they called 'Take the Best' – so called as it simply worked through the cues in order of their predictive validity until one was found that discriminated between two cities, and then responded accordingly. Thus, if the two cities under consideration could not be discriminated on the basis of the most diagnostic cue (e.g. whether it is the state capital or not) then the search through memory continues. The search for discriminatory cue values proceeds in order of their relative diagnosticity until a cue is found that discriminates the two cities (e.g. one has an inter-city rail station and the other does not), whereupon information retrieval is stopped and the judgement made according to this single cue.

Gigerenzer and Goldstein compared simple heuristics such as 'Take the Best' with other decision rules that integrate multiple bits of information (such as multiple regression). They modelled the effect of limited knowledge by simulating six classes of subjects who knew varying proportions of the cue values associated with the cities. Surprisingly, they found that 'Take the Best' did as well as any of the other algorithms and considerably better than some. As it only uses one piece of information it would be much faster than any process that retrieves multiple bits of information and attempts integration of the information. The result is important for demonstrating that, although adherence to normative rules may be sufficient for good judgement, it is not necessary.

Further demonstrations of the efficacy of fast and frugal heuristics have studied binary decisions in a wide range of types of knowledge environment (e.g. which professor has the highest salary? Which US city has more homeless people?). These studies extended the application beyond choice to value estimation, categorization, and memory (Gigerenzer *et al.*, 1999). Goldstein and Gigerenzer (2002) asked American and German students which is bigger: San Antonio or San Diego? While 62 per cent of the Americans correctly named San Diego, 100 per cent of the German students were correct. The Germans were applying a **recognition heuristic** – if you recognize one and not the other, pick the city you have heard of. As you usually hear about the bigger cities of foreign countries before the smaller ones, this will be a pretty good cue. Because the Americans had heard of both cities they couldn't apply this cue and had to rely on other, apparently less valid, cues. In the same way, for city-size decisions, American students were slightly more accurate about German cities than American cities. A large number of subsequent studies confirm that ignorance can even sometimes be helpful because simple mental heuristics *can* exploit the structure of information in the environment to make good inferences (see Gigerenzer and Goldstein, 2011). As a consequence of such results, we might question the present pre-eminent status of normative rules for defining rationality and for serving as a benchmark for assessing human judgement.

SUMMARY OF SECTION 6

- Human decision making may employ fast and frugal heuristics – simple rules that yield quick decisions yet can be highly accurate in certain natural environments.

7 CONCLUSION

The idea that people don't decide as they should was appreciated by psychologists very early on. In a seminal paper, which effectively introduced the study of decision making to psychology, Edwards (1954, p.382) wrote: 'It is easy for a psychologist to point out that an economic man … is very unlike a real man.' Yet for economists this disparity is less clear. As Lopes (a psychologist) put it: 'Economics considers itself a normative science, the very term an oxymoron of ought and is' (1994, p.222). Psychologists and economists think rather differently about the behavioural research exploring decision making (cf. Hertwig and Ortmann, 2001; Lopes, 1994). To psychologists, it is evident that people cannot conceivably represent all the relevant information that normative models require for judgement and decision: 'Who could design a brain that could perform the way this model mandates? Every single one of us would have to know and understand everything completely, and at once' (Daniel Kahneman, quoted by Bernstein, 1996).

While it may be a tad optimistic to presume that the social sciences are on the verge of reconciliation and consensus on this subject, psychology has, since the 1950s, made enormous progress in establishing that actual human decision making cannot be satisfactorily characterized in the idealized way that mathematics and economics have assumed. Moreover, alternative descriptive theories that account for the discrepancies are emerging. Perhaps the best recent piece of evidence for that claim is that a psychologist – Daniel Kahneman – shared the 2002 Nobel Prize for Economics 'for having integrated insights from psychological research into economic science, especially concerning human judgement and decision making under uncertainty' (Nobel Foundation, 2002).

Does the rejection of normative theory as a model for human judgement and choice imply that judgement and choice must be poor or even 'irrational'? No. Although the evidence that people do not perform ideally is clear, any reasonable standards of rationality must surely accept that the computational requirements of normative models are beyond the capacity of a human brain; nevertheless, such bounded rationality (cf. Simon, 1956) does not imply irrationality.

In my (admittedly fallible) judgement the issues surrounding the nature and evaluation of human judgement and decision making are profound and will not be resolved easily or in the near future. To make further progress, we need studies that do more than merely knock down the straw man defined by normative models. Among the many questions that arise, two broad issues can be framed: first, how is it that we are as competent as we evidently are? Second, what can we do about how incompetent we evidently are? Quite how it is that people perform as effectively as they do by applying non-normative mental strategies to the limited information that they can process – and how we might learn to improve our decision making – remain to be explored and explained.

FURTHER READING

Hardman, D.K. (2009) *Judgment and Decision Making: Psychological Perspectives*, Chichester, John Wiley & Sons.

Koehler, D. and Harvey, N. (eds.) (2004) *Blackwell Handbook of Judgment and Decision Making*, Oxford, Blackwell Publishing. Due to limited space some important issues and many interesting developments could not be included here. The roles of emotions, group and inter-personal judgement, and how choices vary over time are not covered. For these issues and more on the topics I have covered I recommend David Hardman's textbook *Judgment and Decision Making* and the *Blackwell Handbook of Judgment and Decision Making*. Both provide good coverage of the topics omitted here and much more besides.

REFERENCES

Allais, M. (1953) 'Le comportement de l'homme rationnel devant le risque, critique des posulats et axioms de l'école Américaine', *Econometrica*, vol.21, pp.503–46.

Allais, M. (1979) 'The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American school', in Allais, M. and Hagen, O. (eds.) *Expected Utility Hypothesis and the Allais Paradox*, Dordrecht, Reidel.

Allais, M. and Hagen, O. (eds.) (1979) *Expected Utility Hypothesis and the Allais Paradox*, Dordrecht, Reidel.

Bar-Hillel, M. (1980) 'The base-rate fallacy in probability judgements', *Acta Psychologica*, vol.44, pp.211–33.

Bernstein, P.L. (1996) *Against the Gods: The Remarkable Story of Risk*, New York, Wiley.

Brunswik, E. (1943) 'Organismic achievement and environmental probability', *Psychological Review*, vol.50, pp.255–72.

Brunswik, E. (1955) 'Representative design and probabilistic theory in a functional psychology', *Psychological Review*, vol.62, pp.193–217.

Budescu, D., Wallsten, T.S., and Au, W.T. (1997) 'On the importance of random error in the study of probabilistic judgement. Part II: Applying the stochastic judgement model to detect systematic trends', *Journal of Behavioral Decision Making*, vol.10, pp.173–88.

Camerer, C.F. (2000) 'Prospect theory in the wild: evidence from the field', in Kahneman, D. and Tversky, A. (eds.) *Choices, Values, and Frames*, Cambridge, UK, Cambridge University Press.

Camerer, C.F., Babcock, L., Loewenstein, G. and Thaler, R. (1997) 'Labor supply of New York City cab drivers: one day at a time', *Quarterly Journal of Economics*, vol.112, pp.407–41.

Christensen-Szalanski, J.J.J. and Bushyhead, J.B. (1981) 'Physicians' use of probabilistic information in a real clinical setting', *Journal of Experimental Psychology, Human Perception and Performance*, vol.7, pp.928–35.

Cohen, L.J. (1981) 'Can human irrationality be experimentally demonstrated?' *The Behavioral and Brain Sciences*, vol.4, pp.317–70.

DuCharme, W.M. and Peterson, C.R. (1968) 'Intuitive inference about normally distributed populations', *Journal of Experimental Psychology*, vol.78, pp.269–75.

Edwards, W. (1954) 'The theory of decision making', *Psychological Bulletin*, vol.41, pp.380–417.

Edwards, W. (1955) 'The prediction of decisions among bets', *Journal of Experimental Psychology*, vol.50, pp.201–14.

Edwards, W. (1968) 'Conservatism in human information processing', in Kleinmuntz, B. (ed.) *Formal Representation of Human Judgment*, New York, Wiley.

Edwards, W. (1992) 'Toward the demise of economic man and woman: bottom lines from Santa Cruz', in Edwards, W. (ed.) *Utility Theories: Measurements and Applications*, Dordrecht, Kluwer.

Erev, I., Wallsten, T.S., and Budescu, D.V. (1994) 'Simultaneous over- and underconfidence: the role of error in judgement processes', *Psychological Review*, vol.101, pp.519–28.

Evans, J.St.B.T., Handley, S.J., Perham, N., Over, D.E., and Thompson, V.A. (2000) 'Frequency versus probability formats in statistical word problems', *Cognition*, vol.77, pp.197–213.

Gigerenzer, G. (1994) 'Why the distinction between single event probabilities and frequencies is important for psychology and vice-versa', in Wright, G. and Ayton, P. (eds.) *Subjective Probability*, Chichester, Wiley.

Gigerenzer, G. and Goldstein, D.G. (1996) 'Reasoning the fast and frugal way: models of bounded rationality', *Psychological Review*, vol.103, pp.650–69.

Gigerenzer, G. and Goldstein, D.G. (2011) 'The recognition heuristic: a decade of research', *Judgment and Decision Making*, vol.6, pp.100–21.

Gigerenzer, G., Hell, W., and Blank, H. (1988) 'Presentation and content: the use of base rates as a continuous variable', *Journal of Experimental Psychology: Human Perception and Performance*, vol.14, pp.513–25.

Gigerenzer, G., Hoffrage, U., and Kleinbölting, H. (1991) 'Probabilistic mental models: a Brunswikian theory of confidence', *Psychological Review*, vol.98, pp.506–28.

Gigerenzer, G., Todd, P.M., and The ABC Research Group (1999) *Simple Heuristics that Make us Smart*, Oxford, Oxford University Press.

Girotto, V. and Gonzales, M. (2002) 'Chances and frequencies in probabilistic reasoning: rejoinder to Hoffrage, Gigerenzer, Krauss, and Martignon', *Cognition*, vol.84, pp.353–9.

Goldstein, D.G. and Gigerenzer, G. (2002) 'Models of ecological rationality: the recognition heuristic', *Psychological Review*, vol.109, pp.75–90.

Grether, D.M. and Plott, C.R. (1979) 'Economic theory of choice and the preference reversal phenomenon', *The American Economic Review*, vol.69, pp.623–38.

Hadar, L. and Fox, C.R. (2009) 'Information asymmetry in decision from description versus decision from experience', *Judgment & Decision Making*, pp.317–25.

Harvey, N. (1997) 'Confidence in judgement', *Trends in Cognitive Sciences*, vol.1, pp.78–82.

Hertwig, R. and Ortmann, A. (2001) 'Experimental practices in economics: a methodological challenge for psychologists?' *Behavioral and Brain Sciences*, vol.24, p.383.

Hertwig, R., Barron, G., Weber, E.U., and Erev, I. (2004) 'Decisions from experience and the effect of rare events in risky choice', *Psychological Science*, vol.15, pp.534–9.

Hoffrage, U., Gigerenzer, G., Krauss, S., and Martignon, L. (2002) 'Representation facilitates reasoning: what natural frequencies are and what they are not', *Cognition*, vol.84, pp.343–52.

Hsee, C.K. (1998) 'Less is better; when low-value options are valued more highly than high-value options', *Journal of Behavioral Decision Making*, vol.11, pp.107–21.

Juslin, P. (1994) 'The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items', *Organizational Behavior and Human Decision Processes*, vol.57, pp.226–46.

Juslin, P., Wennerholm, P., and Olsson, H. (1999) 'Format dependence in subjective probability calibration', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.28, pp.1038–52.

Juslin, P., Winman, A., and Hansson, P. (2007) 'The naïve intuitive statistician: a naïve sampling model of intuitive confidence intervals', *Psychological Review*, vol.114, pp.678–703.

Juslin, P., Winman, A., and Olsson, H. (2000) 'Naive empiricism and dogmatism in confidence research: a critical examination of the hard-easy effect', *Psychological Review*, vol.107, pp.384–96.

Kahneman, D. and Tversky, A. (1972) 'Subjective probability: a judgement of representativeness', *Cognitive Psychology*, vol.3, pp.430–54.

Kahneman, D. and Tversky, A. (1973) 'On the psychology of prediction', *Psychological Review*, vol.80, pp.237–51.

Kahneman, D. and Tversky, A. (1979) 'Prospect theory: an analysis of decision making under risk', *Econometrica*, vol.47, pp.263–91.

Kahneman, D. and Tversky, A. (1982) 'On the study of statistical intuitions', in Kahneman, D., Slovic, P., and Tversky, A. (eds.) *Judgement under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press.

Kahneman, D. and Tversky, A. (1996) 'On the reality of cognitive illusions: a reply to Gigerenzer's critique', *Psychological Review*, vol.103, pp.582–91.

Kahneman, D., Knetsch, J.L., and Thaler, R. (1990) 'Experimental tests of the endowment effect and the Coase theorem', *Journal of Political Economy*, vol.98, pp.1325–48.

Kahneman, D., Slovic, P., and Tversky, A. (eds.) (1982) *Judgement under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press.

Keller, L.R. (1985) 'The effects of problem representation on the sure-thing and substitution principles', *Management Science*, vol.31, pp.738–51.

Keren, G.B. (1987) 'Facing uncertainty in the game of bridge: a calibration study', *Organizational Behavior and Human Decision Processes*, vol.39, pp.98–114.

Klayman, J., Soll, J.B., Gonzalez-Vallejo, C., and Barlas, S. (1999) 'Overconfidence: it depends on how, what, and whom you ask', *Organizational Behavior and Human Decision Processes*, vol.79, pp.216–47.

Knetsch, J.L. and Sinden, J.A. (1984) 'Willingness to pay and compensation demanded: experimental evidence of an unexpected disparity in measures of value', *Quarterly Journal of Economics*, vol.99, pp.507–21.

Koehler, J.J. (1995) 'The base-rate fallacy reconsidered – descriptive, normative, and methodological challenges', *Behavioral and Brain Sciences*, vol.19, pp.1–55.

Kunda, Z. (1990) 'The case for motivated reasoning', *Psychological Bulletin*, vol.108, pp.480–98.

Lichtenstein, S. and Slovic, P. (1973) 'Response-induced reversals of preference in gambling: an extended replication in Las Vegas', *Journal of Experimental Psychology*, vol.101, pp.16–20.

Lichtenstein, S., Slovic, P., and Zink, D. (1969) 'Effect of instruction in expected value on optimality of gambling decisions', *Journal of Experimental Psychology*, vol.79, pp.236–40.

Lopes, L.L. (1994) 'Psychology and economics: perspectives on risk, cooperation, and the marketplace', *Annual Review of Psychology*, vol.45, pp.197–227.

McClelland, A.G.R. and Bolger, F. (1994) 'The calibration of subjective probabilities: theories and models 1980–1994', in Wright, G. and Ayton, P. (eds.) *Subjective Probability*, New York, Wiley.

McKenzie, C.R.M., Liersch, M.J., and Yaniv, I. (2008) 'Overconfidence in interval estimates: what does expertise buy you?' *Organizational Behavior and Human Decision Processes*, vol.107, pp179–91.

Murphy, A.H. and Winkler, R.L. (1984) 'Probability forecasting in meteorology', *Journal of the American Statistical Association*, vol.79, pp.489–500.

Nisbett, R. and Ross, L. (1980) *Human Inference: Strategies and Shortcomings*, Englewood Cliffs, NJ, Prentice Hall.

Nobel Foundation (2002) *The Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel*, http://www.nobel.se/economics/laureates/2002/index.html (accessed 27 July 2004).

Peterson, C.R., Schneider, R.J., and Miller, A.J. (1965) 'Sample size and the revision of subjective probability', *Journal of Experimental Psychology*, vol.69, pp.522–7.

Phillips, L.D. (1984) 'A theory of requisite decision-models', *Acta Psychologica*, vol.56, pp.29–48.

Phillips, L.D. (1987) 'On the adequacy of judgmental probability forecasts', in Wright, G. and Ayton, P. (eds.) *Judgemental Forecasting*, Chichester, Wiley.

Phillips, L.D. (1989) 'Decision analysis in the 1990s', in Shahini, A. and Stainton, R. (eds.) *Tutorial Papers in Operational Research 1989*, Birmingham, Operational Research Society.

Phillips, L.D. and Edwards, W. (1966) 'Conservatism in simple probability inference tasks', *Journal of Experimental Psychology*, vol.72, pp.346–57.

Pitz, G.F., Downing, L., and Rheinold, H. (1967) 'Sequential effects in the revision of subjective probabilities', *Canadian Journal of Psychology*, vol.21, pp.381–93.

Raiffa, H. (1968) *Decision Analysis*, Reading, MA, Addison Wesley.

Rakow, T. and Newell, B.R. (2010) 'Degrees of uncertainty: an overview and framework for future research on experience-based choice', *Journal of Behavioral Decision Making*, vol.23, pp.1–14.

Rakow, T., Demes, K.A., and Newell, B.R. (2008) 'Biased samples not mode of presentation: re-examining the apparent underweighting of rare events in experience-based choice', *Organizational Behavior and Human Decision Processes*, vol.106, pp.168–79.

Samuelson, P.A. (1963) 'Risk and uncertainty: a fallacy of large numbers', *Scientia*, vol.98, pp.108–13.

Samuelson, W. and Zeckhauser, R. (1988) 'Status quo bias in decision making', *Journal of Risk & Uncertainty*, vol.1, pp.7–59.

Savage, L. (1954) *The Foundations of Statistics*, New York, Wiley.

Schlaifer, R. (1969) *Analysis of Decisions under Uncertainty*, New York, McGraw Hill.

Shafir, E. (1993) 'Choosing versus rejecting: why some options are both better and worse than others', *Memory and Cognition*, vol.21, pp.546–56.

Shafir, E., Simonson, I., and Tversky, A. (1993) 'Reason-based choice', *Cognition*, vol.49, pp.11–36.

Shafir, S. (1994) 'Intransitivity of preferences in honeybees: support for "comparative" evaluation of foraging options', *Animal Behaviour*, vol.48, pp.55–67.

Simon, H.A. (1956) 'Rational choice and the structure of the environment', *Psychological Review*, vol.63, pp.129–38.

Slovic, P. (1975) 'Choice between equally valued alternatives', *Journal of Experimental Psychology: Human Perception and Performance*, vol.1, pp.280–7.

Slovic, P. (1995) 'The construction of preference', *American Psychologist*, vol.50, pp.364–71.

Slovic, P. and Lichtenstein, S. (1968) 'Relative importance of probabilities and payoffs in risk taking', *Journal of Experimental Psychology Monograph*, vol.78, pp.1–18.

Slovic, P. and Tversky, A. (1974) 'Who accepts Savage's axiom?' *Behavioral Science*, vol.19, pp.368–73.

Slovic, P., Fischhoff, B., and Lichtenstein, S. (1977) 'Behavioral decision theory', *Annual Review of Psychology*, vol.28, pp.1–39.

Soll, J.B. and Klayman, J. (2004) 'Overconfidence in interval estimates', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.30, pp.299–314.

Tversky, A. (1969) 'Intransitivity of preferences', *Psychological Review*, vol.76, pp.31–48.

Tversky, A. and Kahneman, D. (1974) 'Judgement under uncertainty: heuristics and biases', *Science*, vol.185, pp.1124–31.

Tversky, A. and Kahneman, D. (1981) 'The framing of decisions and the psychology of choice', *Science*, vol.211, pp.453–8.

Tversky, A. and Kahneman, D. (1983) 'Extensional versus intuitive reasoning: the conjunction fallacy in probability judgement', *Psychological Review*, vol.90, pp.293–315.

Tversky, A. and Kahneman, D. (1986) 'Rational choice and the framing of decisions', *Journal of Business*, vol.59, S251–S278.

Tversky, A., Sattath, S., and Slovic, P. (1988) 'Contingent weighting in judgement and choice', *Psychological Review*, vol.95, pp.371–84.

Tversky, A., Slovic, P., and Kahneman, D. (1990) 'The causes of preference reversal', *American Economic Review*, vol.80, pp.204–17.

Ungemach, C., Chater, N., and Stewart, N. (2009) 'Are probabilities overweighted or underweighted, when rare outcomes are experienced (rarely)?' *Psychological Science*, vol.20, pp.473–9.

von Neumann, J. and Morgenstern, O. (1944) *Theory of Games and Economic Behavior*, Princeton, Princeton University Press.

von Winterfeldt, D. and Edwards, W. (1986) *Decision Analysis and Behavioral Research*, Cambridge, Cambridge University Press.

Wagenaar, W.A. and Keren, G.B. (1986) 'Does the expert know? The reliability of predictions and confidence ratings of experts', in Hollnagel, E., Mancini, G., and Woods, D.D. (eds.) *Intelligent Decision Support in Process Environments*, Berlin, Springer-Verlag.

Watson, S.R. and Buede, D.M. (1987) *Decision Synthesis: The Principles and Practice of Decision Analysis*, Cambridge, Cambridge University Press.

Wilson, T.D., Houston, C.E., Etling, K.M., and Brekke, N. (1996) 'A new look at anchoring effects: basic anchoring and its antecedents', *Journal of Experimental Psychology: General*, vol.125, pp.387–402.

Winkler, R.L. and Murphy, A.M. (1973) 'Experiments in the laboratory and the real world', *Organizational Behavior and Human Performance*, vol.20, pp.252–70.

Winman, A., Hansson, P., and Juslin, P. (2004) 'Subjective probability intervals: how to cure overconfidence by interval evaluation', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.30, pp.1167–75.

Yaniv, I. and Foster, D.P. (1997) 'Precision and accuracy of judgmental estimation', *Journal of Behavioral Decision Making*, vol.10, pp.21–32.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

CHAPTER 12

REASONING

Mike Oaksford

1 INTRODUCTION

Suppose a friend tells you that ‘if John finds out, he will be furious’ (i). You later discover that ‘John found out’ (ii). If you conclude that ‘John was furious’ (iii), then you have engaged in a passage of reasoning: you have inferred a conclusion (iii) from some premises, i.e. (i) and (ii). Reasoning has been studied since the time of the ancient Greeks when Aristotle suggested that reasoning is one of the abilities that mark us out from the other animals. Only we use language and only we reason. Therefore only we possess minds capable of rational thought. Aristotle also developed the first system of logic. That is, he produced a set of rules by which to judge whether certain passages of reasoning, like the one above, were valid, i.e. for telling whether the conclusion (iii) really does follow from the premises, (i) and (ii). This is one of the sources of a strong line adopted by some researchers in this area: reasoning is the process of applying logical laws. This strong line also emerges in the foundations of modern logic in Boole’s *The Laws of Thought* (1854). In that book, Boole describes a set of rules that determine how we can draw inferences from statements like ‘if … then’ (as in example (i)). The book’s title clearly reveals that the author’s intention was to describe the laws that determine human reasoning. More recently, Piaget placed the ability to reason according to logical rules at the pinnacle of his stage theory of cognitive development, i.e. at the formal operational stage (Inhelder and Piaget, 1958). One of our main concerns in this chapter will be to determine whether logic provides a good model of human reasoning. This does not mean that we should expect people to reason perfectly logically. Although we may all be capable of logical performance, perfect performance may, for example, take up too much memory. In which case, we might expect some errors to emerge and, as we will see, such errors have been observed.

Another theme that will emerge is whether logic is appropriate to describe real human reasoning at all. The full title of Boole’s (1854) book was *An Investigation of the Laws of Thought, on Which are Founded the Mathematical Theories of Logic and Probabilities*. Almost half of the book was devoted to probability theory, which Boole thought may provide a better theory of everyday reasoning. As we will see, some researchers do not view people’s reasoning behaviour as error prone but logical. Rather they view it as relatively error free but probabilistic. This difference emerges as a result of assigning differing meanings to the important words that figure in a passage of reasoning. In this chapter, we will concentrate heavily on the words 'if . . . then'. According to some researchers, 'if . . . then' should be interpreted logically, in terms of the conditions under which it is true or false. For example, (i) is false if John does not get furious when he finds out; otherwise it is true. Other researchers view (i) as describing a causal relation between two events, i.e. John finding out causes him to get furious. On this view, (i) should be interpreted probabilistically, in terms of the evidence one might have for John getting furious given he finds out. Of course, different people may know different things about John, which may lead them to different evaluations of the probability that he gets furious when he finds out. This could lead to individual differences in reasoning, and as we will see, such differences have also been observed.

A further theme of this chapter arises from the fact that the experimental work on human reasoning shows that according to logic, people do make many errors. If being logical is what we mean by being rational then these results may have some serious consequences. For example, in law people can only be held responsible for their actions if they can rationally evaluate the consequences of those actions. But if people are not rational, then how can society hold them responsible for what they do? Moreover, where is the boundary between sanity and insanity to be drawn? A sane person is one who responds rationally to the world and to other people. But if most normal adults are irrational then who is sane and who is insane? In evaluating psychological theories in this area, it will therefore be important for us to pay close attention to what they have to say about human rationality.

The first issue we address is the sheer ubiquity of human reasoning. Reasoning may seem like an activity we rarely, if ever, engage in. However, most of our common-sense psychological explanations of each other's behaviour assume that we are reasoning all the time.

1.1 Reasoning in everyday life

People are so dependent on reasoning processes that they tend to go unnoticed. Nonetheless, it is simple to show that a great deal of human behaviour depends on reasoning processes. Suppose you see your neighbour's wife arriving home. She passes her garage and she sees that her husband's car is on the drive. When she reaches the door, instead of taking out her key and opening the door as she has done every night for the last 20 years, she rings the doorbell. Why your neighbour broke her habitual pattern of behaviour can be explained in an instant:

> *She saw that the car was there. She knows that for it to be there, someone must have driven it from town where her husband dropped her off in the morning. Because she knows that only her husband has the keys, she infers that her husband drove it. She further infers that if the car is on the drive then her husband is in the house and hence he can open the door when she rings the bell. Consequently, rather than take out her key, she rings the doorbell.*

Two pieces of information combine in the first inference. One comes directly from perception, i.e. *the car was there*, and the other from prior knowledge, i.e. *if the car was there then someone drove it*. This information is combined in an *inference* to yield new information: someone drove the car.

The given information can be regarded as the *premises* and the new information as the *conclusion* of a passage of reasoning. The subsequent steps that lead her to the final conclusion that she should ring the doorbell can all be characterized in the same way. So it

(1)	*If the car is there someone drove it.* The car is there.	(Premises)
	Therefore, Someone drove the car.	(Conclusion)

would seem that even the most mundane passage of human behaviour involves complex reasoning processes that require using given information (premises) to infer new information (conclusions).

The question we now need to address is how do we know when we can draw a conclusion from some set of premises? For example, the conclusion in (1) could be replaced with 'Her partner had a cream tea'. This would be new information too, but it does not seem to be related to the premises in the right kind of way: believing the premises in (1) in no way compels you to believe this conclusion. Trying to describe this relationship between premises and conclusions is the core of characterizing *logical* or *deductive* reasoning. The idea, as we saw in Aristotle and Boole, is to provide some rules to determine when a conclusion does and does not follow.

2 DEDUCTIVE REASONING AND LOGIC

Logic provides an account of the relationship between the premises and the conclusion of a deductive argument. Logic starts from the idea that sentences are made up of two very general building blocks. First there are descriptive clauses that say something true or false about the world, e.g. 'John has a runny nose' or 'John has a cold'. These sentences are said to describe *propositions*, i.e. they describe how the world may or may not be. Second, there are structure-building words that allow us to combine propositions to produce more complex sentences, e.g. 'John has a runny nose *and* John has a cold' or '*if* John has a runny nose, *then* he has a cold'. Along with *and* and *if . . . then*, other structure-building words include *or* and *not*. Collectively these words are called *connectives* because they connect two simpler sentences together. The most important of these is *if p then q*, which is called the *conditional*. The *p* clause is called the *antecedent* and the *q* clause is called the *consequent*.

Propositions are simply true or false depending on whether what they say holds in the world or not, so the sentence 'John has a runny nose' is true if and only if John has a runny nose, otherwise it is false. However, how do we determine whether the more complex sentences made using the structure-building words or connectives are true or false? Well, *not* is particularly simple because all it does is reverse the truth value of a proposition. So if it is true that 'John has a runny nose' then 'John does not have a runny nose' must be false. Conversely, if it is false that 'John has a runny nose' then 'John does not have a runny nose' must be true. What about the other connectives? Notice that the other three connectives all connect two propositions, e.g. 'John has a runny nose *and* John has a cold'. Each proposition making up this complex sentence can be true or false. Thus there are four possibilities, letting p = 'John has a runny nose' and q = 'John has a cold': either p is true and q is true; or p is true and q is false; or p is false and q is true; or p is false and q is false. In only one of these possibilities would we intuitively say that 'John has a runny nose *and* John has a cold' is true, i.e. when p is true and q is true, otherwise this complex sentence is false. So we can regard the connective '*and*' as mapping pairs of truth values on to a truth value. When both p and q are true, then p *and* q is true; for all other pairs it is false. We can show this mapping in what is called a truth table (see Table 12.1).

With these truth tables it is easy to show what we said about the relation between the premises and the conclusion of a deductive argument. We said that if you believed the premises of a deductive argument then somehow you *must* believe the conclusion. This means, for example, that if it is true that 'if John has a runny nose then he has a cold’ and it is true that ‘John has a runny nose’, then it has to be true that ‘John has a cold’. The truth tables in Table 12.1 allow us to show this. Again, let p = ‘John has a runny nose’ and q = ‘John has a cold’. If we look at Table 12.1, then we see that *if p then q* and p are only both true when q is also true (first row of the table). So if both the premises, *if p then q* and p are true, then the conclusion, q, must be true; there is no other possibility! We describe this by saying that the inference is *logically valid*. This inference is also referred to as *modus ponens*, or MP for short, and it typically appears in logic textbooks as:

TABLE 12.1 Truth tables. Truth tables for the logical connectives *and*, *if … then*, and *or*. The two leftmost columns show the four possible combinations of truth values for the propositions *p* and *q*. They also figure as one of the premises and the conclusion of an inference involving the connectives. The columns for each connective show the truth value each particular pair is mapped onto by that connective.

p	q	p and q	if p then q	p or q
true	*true*	*true*	*true*	*true*
true	*false*	*false*	*false*	*true*
false	*true*	*false*	*true*	*true*
false	*false*	*false*	*true*	*false*

(2)	MP		*If p then q*	*If John has a runny nose, then he has a cold*
			p	*John has a running nose*
		∴	*p*	*John has a cold*

Another logically valid inference with the conditional is *modus tollens*, or MT for short (3). We can again show this using Table 12.1. Of course, *not-q* is true when q is false. Consequently, *if p then q* and *not-q* are only both true in the last line of the truth table where p is false, and so *not-p* is true. So if these two premises are true then *not-p* must be true also, as again there is no other possibility. So if the conditional is true and John does not have a cold, then John cannot have a runny nose.

(3)	MT		*If p then q*	*If John has a runny nose, then he has a cold*
			not-q	*John does not have a cold*
		∴	*not-p*	*John does not have a runny nose*

There are two other inference patterns investigated in the psychology of reasoning: *affirming the consequent* (AC) and *denying the antecedent* (DA):

(4)	AC		*If p then q*	*If John has a runny nose, then he has a cold*
			q	*John has a cold*
		∴	*p*	*John has a runny nose*

(5)	DA		*If p then q*	*If John has a runny nose, then he has a cold*
			not-q	*John does not have a running nose*
		∴	*not-p*	*John dose not have a cold*

These inferences are *not* logically valid.

✱ ACTIVITY 12.1

See if you can work out why AC and DA are not logically valid from Table 12.1. As a clue, we found that MP and MT were valid because in the only line in the truth table where the premises were both true, the conclusion was true. Is there a line in the truth table (Table 12.1) for which the premises of AC or DA are true but the conclusion is false?

We can now illustrate a critical distinction in logic between *form* and *meaning*. (2) and (3) show the *form* of the logically valid inferences MP and MT. We know they are logically valid because of the *meanings* we have assigned to the connectives in the truth tables in Table 12.1. These tell us that if the premises are true then the conclusion must be true, which is the definition of logical validity. However, to draw the MP or MT inference does not require that we make reference to the truth table, i.e. to the meaning of the conditional. For example, to draw the inference in (1) all you would need to do is to match it to the *formal rule* in (2). That is, (1) is simply a particular instantiation of (2). With (2) we can automatically make the logically valid inference simply by virtue of the *form* of the argument without having to worry about what it *means*. This distinction between *form* and *meaning* is central to two of the theories of reasoning we look at later on. One theory, *mental logic*, argues that we have formal inference rules, like (2) to (5), in our heads, i.e. inferences only rely on *form*. The other, *mental models*, argues that we do something much more like considering the truth tables, i.e. inferences rely on the *meaning* of the connectives.

SUMMARY OF SECTION 2

- Logic provides a set of rules to determine if an inference is logically valid.
- Logic divides words into propositions and structure-building words.
- Propositions are true or false; structure-building words or connectives map pairs of truth values onto a truth value.
- These mappings can be displayed in truth tables, which can be used to determine if an inference is logically valid. If when the premises are true, the conclusion must be true, then the inference is valid.
- This semantic way of establishing validity contrasts with the formal method of using rules of inference (2) and (3).

3 PSYCHOLOGICAL THEORIES OF REASONING

In the psychology of human reasoning there are a wide variety of theoretical approaches. In this section the main theoretical approaches are outlined and discussed. Three theories are introduced: *mental logic*, *mental models*, and the *probabilistic approach*. These are *general* theories of reasoning, i.e. they are intended to apply to most reasoning tasks.

3.1 Mental logic

This group of theories – there are several different versions of this basic account (Braine and O'Brien, 1998; Rips, 1994) – are also known as *formal rule* theories. As the name suggests, these accounts are close in spirit to Piaget, who argued that adult human thought *was* the operation of formal logic (Inhelder and Piaget, 1958). The idea behind these theories is that people possess a system of formal mental logic that contains inference rules like (2) and (3). However, people may not possess all the formal logical rules that are licensed by the truth tables in Table 12.1. Without a particular rule some inferences might be more complicated than others.

3.2 Mental models

Mental models theory (Johnson-Laird, 1983, 2006; Johnson-Laird and Byrne, 1991) shares the intuition with mental logic that people are at least in principle capable of logical reasoning. However, rather than applying formal rules, mental models theory argues that people reason over pictorial representations of what those sentences *mean*. These representations concern the different *possibilities* that a logical expression may allow. This is familiar from our discussion of truth tables. In Table 12.1, each row corresponds to a different possibility depending on whether *p* and *q* are true or false, which leads to four possibilities. One way of thinking about the connectives is that they exclude different possibilities. So for example, if *if p then q* is true then the possibility where *p* is true and *q* is false is excluded, leaving three other possibilities. As we know (see Chapter 9), working memory has limited capacity. This means that people may not represent all these possibilities at once when interpreting a connective. Rather, there may be a preferred initial representation. This is the core of the mental models theory. In this account each possibility is referred to as a 'mental model'. How people manipulate these mental models explains their reasoning performance.

3.3 The probabilistic approach

According to the probabilistic approach, the goal of a psychological theory of reasoning is to account for people's everyday inferences (Oaksford and Chater, 1994, 1998, 2007, 2009). According to this view, logic is not the right theory of how people should make these inferences. For example, according to logic a good reason to believe that *if John has a runny nose then he has a cold* is that you do not believe that *John has a runny nose*. This is because, according to logic, a conditional is true when its antecedent is false, i.e. lines 3 and 4 of Table 12.1. However, not believing that *John has a runny nose* does not seem to be sufficient grounds to believe that this conditional is true. What would appear to be required is the belief that John having a runny nose makes it very likely that he has a cold. This involves assessing the *conditional probability* that John has a cold given he has a runny nose. So if you have noticed John having a runny nose on, say, 100 occasions, 95 of which involved him having a cold, then the relevant conditional probability is 0.95, i.e. $P(\text{John has a cold} \mid \text{John has a runny nose}) = 0.95$ (the bar '|' = given). This is then your degree of belief in the conditional. It is also the probability with which people would be expected to draw the MP inference. If all you know is that John has a runny nose, then your degree of belief in this rule indicates that there is a 95 per cent chance that he has a cold. According to the probabilistic approach, most inferences are *uncertain* because the conditional rules on which they are based describe the real world in which logical certainty is a rare commodity. This account of reasoning performance therefore suggests replacing logic with probability theory as the account of the inferences people should make.

The contrast between the first two general theories of reasoning and the last is important because the concept of what it is to be rational fundamentally changes. In the first two theories, rationality is still defined as reasoning according to logic. However, people make mistakes mainly because of the limited nature of short-term memory. In contrast, according to the probabilistic approach, probability theory replaces logic as the criterion of what is rational.

Assessing theories in any area of science depends first and foremost on the ability to explain the data. In the next two sections we therefore look at some of the experimental results on the two principal reasoning tasks that have been used to assess human reasoning performance: the conditional inference task (Evans, 1977; Taplin, 1971) and the Wason Selection Task (Wason, 1968).

SUMMARY OF SECTION 3

In this section we have outlined the three main general theories of reasoning.

- In *mental logic theories*, people use formal rules of inference but people do not possess all possible rules. This means that some inferences will be more difficult.
- In *mental models theory*, people represent the true possibilities licensed by a connective. However, they may represent only one of these possibilities, which could be the source of errors.
- In *the probabilistic approach*, people are not trying to draw deductive inferences. The degree to which inferences are endorsed depends on the appropriate conditional probability.

4 CONDITIONAL INFERENCE

In a conditional inference task (Evans, 1977; Taplin, 1971) participants are presented with a conditional sentence as in our example, *if John has a runny nose, then he has a cold* (the conditional premise), and various facts relating to the antecedent or consequent of the sentence (the categorical premise), e.g. *John has a runny nose* (MP), *John does not have a runny nose* (DA), *John has a cold* (AC), *Johnny does not have a cold* (MT). From these pairs of premises, logic dictates that various inferences introduced in Section 2 should be made or withheld. We now review the main findings on conditional inference. For each finding we discuss how the main theories explain the results.

4.1 The abstract conditional inference task

Typically these reasoning tasks were conducted using abstract *alphanumeric* stimuli, i.e. letters and numbers. The idea was to try and rule out any effects of prior knowledge so as to investigate reasoning 'in the raw'. A similar motivation for studying recall of meaningless letter triples, e.g. UVJ, was originally provided in memory research (Ebbinghaus, 1885). In both cases, the idea was to investigate the basic operating characteristics of the cognitive system. Of course, this may just be throwing out the baby with the bath water because the memory and reasoning systems may operate very differently when prior knowledge is relevant, which is surely the norm.

In the abstract conditional inference task participants are told to assume that, for example, the premises *if A then 2* and *there is an A* are true and are asked whether they can conclude that *there is a 2* (MP). In a summary of 65 of these experiments, on average 97 per cent of participants endorse MP, 55 per cent endorse DA, 63 per cent endorse AC, and 72 per cent endorse MT (Schroyens and Schaeken, 2003). If we compare these mean values two at a time, i.e. *pairwise*, then all six comparisons reveal highly significant differences. MP is endorsed more than MT, which is endorsed more than AC, which is endorsed more than DA. This pattern is not consistent with logic, according to which participants should endorse MP and MT equally and should not endorse any other inference.

4.1.1 Mental logic

Mental logicians (Rips, 1994) explain the difference between MP and MT by the proposal that while people possess the MP inference rule ((2) in Section 2) they do not possess the MT inference rule ((3) in Section 2). This means that to draw the MT inference requires indirect and more complex reasoning. The MT inference is, *If John has a runny nose, then he has a cold*, *He does not have a cold*, therefore, *He does not have a runny nose*. Rips (1994) argues that to reach this conclusion, people assume the contrary of the conclusion, i.e. *John has a runny nose*, and show that a contradiction results. If this assumption is true then it can be combined with the conditional premise to yield the conclusion, *John has a cold*, by MP. But this conclusion contradicts the categorical premise, that *John does not have a cold*. So the assumption must be false, i.e. *John has a runny nose* must be false and so *John does not have a runny nose* must be true! This way of drawing the inference is called *reductio ad absurdum* or RAA. The additional complexity of RAA explains why MT is drawn less often than MP.

Indeed, this strategy may also explain the results for DA and AC (Rips, 1994). Conditionals in natural language can be ambiguous and may sometimes be interpreted as *bi-conditionals*. In terms of truth tables (Table 12.1) this means that the conditional is also true when both antecedent (p) and consequent (q) are false. It also means that *if John has a runny nose, then he has a cold* may be taken to also mean its converse that *if John has a cold, then he has a runny nose*. Those people who interpret the rule this way will make just MP and MT on the standard rule. However, they should also make MP and MT on the converse rule. Thus they will look like they are making DA and AC inferences on the standard rule because DA and AC on *if John has a runny nose, then he has a cold* is equivalent to MT and MP (respectively) on *if John has a cold, then he has a runny nose*. Consequently, those participants who interpret the rule as a bi-conditional should endorse both DA and AC but they should endorse AC more than DA. This corresponds to the pattern of endorsements observed in Schroyens and Schaeken's (2003) summary of the data. So, according to the mental logic theory, people's performance on this task is rational because it is actually logical; it is just that some logical inferences are harder than others given the logical rules we possess.

4.1.2 Mental models

Table 12.2 (a) shows the initial mental model representation for the conditional. It represents the possibility in which p is true and q is true, i.e. it is like the first line in Table 12.1. The three dots (or ellipsis) indicate that there may be other relevant mental models. That is, there are other possibilities that are not currently being considered. These other possibilities are available, perhaps held in some temporary short-term memory store, but they are not explicitly represented. The square brackets indicate that p cannot be paired with any other term. In particular, this captures the fact that p cannot be paired with *not-q* ($\neg q$), i.e. line 2 in Table 12.1, because this is the possibility that the conditional excludes. Table 12.2 (c) shows the initial mental model for the bi-conditional. Both p and q are now in square brackets because neither can be paired with anything else. This rules out the p, *not-q* and the *not-p*, q possibilities. Table 12.2 (b) shows the *fleshed-out* mental model for the conditional, where the other possibilities not excluded by this interpretation are explicitly represented. Table 12.2 (d) shows the same fleshed-out mental model for the bi-conditional. The fact that mental models only represent the true possibilities, and the distinction between initial mental models and fleshed-out mental models, are the primary means by which mental models explains the data on conditional inference. The explanation depends on different subsets of participants adopting these four different interpretations.

If a participant adopts the initial conditional interpretation (Table 12.2 (a)) then they can only draw the MP inference. While the categorical premise of both MP (p) and AC (q) match the model, q may be paired with something other than p. However, p can only be paired with q. However, if participants adopt the fleshed-out conditional interpretation (Table 12.2 (b))

TABLE 12.2 Mental models for the conditional. Initial mental models for the conditional (a) and bi-conditional (c) and once they have been fleshed-out (b) and (d). The inferences that can be drawn in each model are also shown.

$[p]$	q	$[p]$	q	$[p]$	$[q]$	$[p]$	$[q]$
. . .		$\neg p$	q	. . .		$\neg p$	$\neg q$
		$\neg q$	$\neg q$				
(a)		(b)		(c)		(d)	
MP		MP MT		MP AC		MP DA AC MT	

they can now also make the MT inference in addition to MP. Although the categorical premise of all inferences now match the model, only *not-q* ($\neg q$) is constrained to only be paired with *not-p* ($\neg p$). If participants adopt the initial bi-conditional interpretation (Table 12.2 (c)) they can make the MP inference and the AC inference. MP goes through for the same reason as for the initial conditional interpretation. AC goes through because now q can only be paired with p. Finally, if participants adopt the fleshed-out bi-conditional interpretation (Table 12.2 (d)), then they can draw all inferences. This is because the categorical premise of all inferences find a match in the model and each is uniquely paired with only a single item.

Table 12.2 shows that MP can be endorsed in all representations, DA can only be endorsed in one, and AC and MT can both be endorsed in two. So if equal numbers of participants adopted each representation, then MP will be endorsed more than both AC and MT, which will be endorsed in equal proportion, and all will be endorsed more than DA. However, participants endorse more MT than AC inferences. Explaining this result assumes that more participants flesh out the conditional interpretation (Table 12.2 (b)) than endorse the initial bi-conditional interpretation (Table 12.2 (c)).

4.1.3 Probabilistic approach

According to the probabilistic approach people will draw inferences depending on how probable they think the conclusion is given the premises (Oaksford and Chater, 2003a; Oaksford *et al.*, 2000). MP is straightforward. Given the conditional premise, *if John has a runny nose* (r), *then he has a cold* (c), this inference will be drawn in proportion to the *conditional probability* that *John has a cold* given that *he has a runny nose* ($P(c|r)$). So for example, given the categorical premise, *John has a runny nose* (and no other information) the best bet as to the probability that *John has a cold* is the proportion of times that John has a cold when he has a runny nose. In our example above, this was 0.95. Calculating probabilities for the remaining inferences requires the assumption that people possess information about the probability of John having a runny nose ($P(r)$) and the probability that he has a cold ($P(c)$). Consider the AC inference in which you are told that *John has a cold* and are asked whether you would endorse the conclusion that *John has a runny nose*. What you are interested in is the probability that *John has a runny nose* given that *John has a cold* ($P(r|c)$), i.e. the converse of MP. Probability theory

✱ ACTIVITY 12.2

For each of the inferences, work through the mental models in Table 12.2 and be sure that you can tell why each inference is valid or not in each model.

You should find that the process is very like checking the truth table in Activity 12.1.

allows us to calculate this probability using *Bayes' Theorem* (see Chapter 11):

$$P(r\,|\,c)=\frac{P(c\,|\,r)P(r)}{P(c\,|\,r)P(r)+P(c\,|\,\neg r)P(\neg r)}$$

We know that $P(c|r) = 0.95$. Suppose that John rarely has a runny nose and so over the year he has only a 5 per cent chance of having one, so $P(r) = 0.05$ and $P(\neg r) = 0.95$. What about $P(c|\neg r)$, i.e. the probability of *John having a cold* given *he does not have a runny nose*? This is likely to be quite low but not 0. Let us set this to 0.03. $P(r|c)$ is then equal to 0.61, which is the probability that you should draw the AC inference. Probabilities can also be derived in a similar way for DA and MT.

To assess how well this account can explain the standard abstract results, the equations predicting the probabilities with which each inference should be drawn can be *fitted* to the data. This means that values of $P(c|r)$, $P(r)$, and $P(c)$ can be estimated that provide the closest possible predictions for the frequencies with which each inference is endorsed. The probabilistic account provides quite a close fit to the abstract data, predicting the following frequencies of drawing each inference (actual frequencies in brackets): MP = 0.88 (0.97), DA = 0.51 (0.55), AC = 0.68 (0.63), MT = 0.77 (0.72) (Oaksford and Chater, 2003b). The ability to fit the data in this way is not unique to the probabilistic approach. The numbers of participants adopting the conditional or bi-conditional interpretation and the numbers fleshing out or drawing RAA inferences are all free to vary and must be fixed from the data, just as for the probabilities in the probabilistic approach.

4.2 Everyday reasoning and the suppression effect

A general property of everyday inferences is that they can be defeated (Oaksford and Chater, 1998). For example, if you infer that *John has a cold* because *he has a runny nose*, but then discover that *he has hay fever*, your inference is defeated. Because such inferences can be defeated in this way they are called *defeasible* inferences. Or to take another example, we generally believe that *birds fly*, which can be expressed as the conditional *if something is a bird then it flies*. So we might conclude that *Tweety can fly* on learning that *Tweety is a bird*. However, if we then discover the new information that *Tweety is an ostrich*, then this inference is defeated. It thus seems that many of the inferences we draw in everyday life may be non-deductive, i.e. the truth of the premises does not guarantee the truth of the conclusion.

There have been many experiments investigating these aspects of everyday reasoning. They show that the inferences, MP and MT, and the fallacies, DA and AC, can be *suppressed* by providing information about possible defeaters. For example, if you are told that *if the key is turned then the car starts* and that *the key is turned*, you are likely to endorse the MP inference to the conclusion that *the car starts*. However, if you are also told that *the petrol tank is empty*, you are less likely to endorse this conclusion because the car will *not* start if the petrol tank is empty. An empty petrol tank provides an *exception* to the rule. This exception would also mean that you are less likely to endorse MT. If you knew that *the car didn't start* you may not infer that *the key was not turned* because the empty petrol tank may be the cause of the car not starting. These exceptions have been called 'additional antecedents' (Byrne, 1989). Other information can suppress DA and AC. For example, if you are told that *if the key is turned then the car starts* and that *the key is not turned*, you might endorse the DA inference to the conclusion that *the car does not start*. However, if you are also told that *the car was hot-wired*, you may be less likely to endorse this conclusion because the car may start even though the key was not turned because it has been hot-wired. This condition would also mean that you are less likely to endorse AC. If you knew that *the car started* you may not infer that *the key was turned* because the car starting may have been caused by it being hot-wired. These conditions have been called 'alternative antecedents' (Byrne, 1989).

Byrne (1989) demonstrated all these effects by providing participants with explicit rules containing this additional information.

(6) Additional antecedents (MP)

If the key is turned the car starts
If there is fuel in the tank the car starts
The key is turned

The car starts?

(7) Alternative antecedents (AC)

If the key is turned the car starts
If it is hot-wired the car starts
The car starts

The car starts?

The results of Byrne's (1989) Experiment 1 are shown in Figure 12.1. The simple condition did not include any additional or alternative antecedents and reflects the standard pattern of results (see Section 4.1). Figure 12.2 shows clearly that alternative antecedents suppress DA and AC but not MP or MT, whereas additional antecedents suppress MP and MT but not DA or AC.

Similar effects have been demonstrated *without* presenting additional or alternative antecedents explicitly (Cummins *et al.*, 1991; Cummins, 1995). Thus, people seem to automatically retrieve this information from memory to influence their reasoning performance. A range of causal conditionals, like those we have looked at, were pre-tested for the number of additional or alternative antecedents participants could bring to mind. It was shown that the effects of additional or alternative antecedents were *graded*, i.e. the more additional or alternative antecedents a rule allowed, the greater the suppression effects observed. As we will see, this result does not sit well with theories that regard these effects as *all or nothing*. In experiments where participants *rated* how likely an inference was to go through on a 1 to 7 scale (Cummins *et al.*, 1991; Cummins, 1995), *all or nothing* behaviour suggests that the data should be made up of different proportions of individuals using only 1 (do not endorse) or 7 (endorse). However, in experiments like these the mean percentage of intermediate values used per participant is around 50 per cent (Oaksford *et al.*, 2000, Experiment 3), which is not consistent with all or nothing behaviour.

4.2.1 Mental logic

In Byrne's (1989) experiments (see (6) and (7)) participants were provided with two conditional rules with the same consequents. So for additional antecedents (6) the key needs to be turned *and* there needs to be fuel in the tank for the car to start. However, for alternative antecedents (7) either turning the key *or* hot-wiring will start the car. This means that people

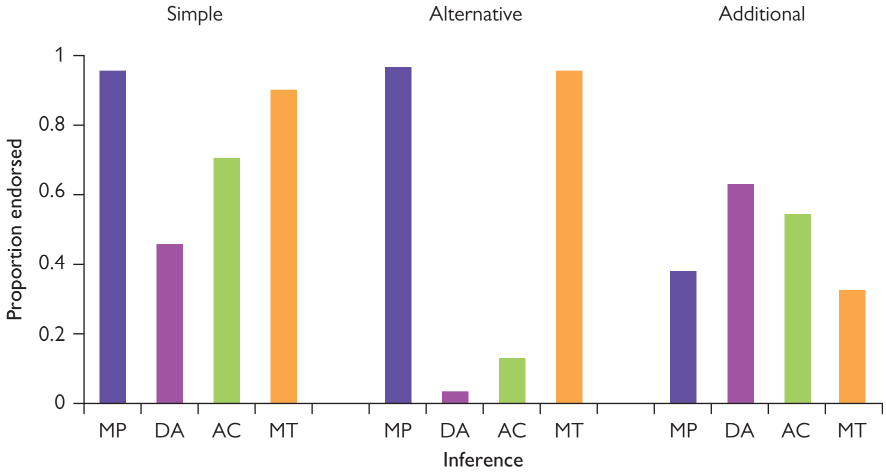


FIGURE 12.1 Suppression effects. *Source: Byrne (1989), Table 1*

✱ ACTIVITY 12.3

For (7′) see if you can work out why DA and AC should be suppressed. A clue for the DA inference is that *not* (*the key is turned* **or** *the car is hot-wired*) is equivalent to *the key is not turned* **and** *the car is not hot-wired*.

may represent these premises as single rules but with complex antecedents:

> *(6′) If the key is turned* **and** *there is fuel in the tank, then the car starts*
> *(7′) If the key is turned* **or** *it is hot-wired, then the car starts*

Let us look at each inference in turn for (6′).

MP Inference: The categorical premise *the key is turned* does *not* satisfy the antecedent, which is only true if *the key is turned* **and** *there is fuel in the tank* (see Table 12.1). Consequently, *logically* this single categorical premise does not licence the conclusion that *the car starts* and so the MP inference should be suppressed.

MT Inference: This leads to the negation of the antecedent, i.e. (*not the key is turned* **and** *there is fuel in the tank*), which is equivalent to *the key is not turned* **or** *there is no fuel in the tank*. Consequently, *logically* you cannot infer that *the key was not turned* because there could be no fuel in the tank, and so the MT inference should be suppressed. Similar reasoning applies to (7′) for the DA and the AC inferences.

Mental logic suggests that reasoning behaviour is all or nothing on this task; this is hard to reconcile with Cummins' (1995) data, which revealed graded effects. Mental logicians have responded by arguing for a clear distinction between deductive and non-deductive inference (Rips, 2001, 2002). The strength of a non-deductive inference can be thought of as how probable the conclusion is given the premises, as in the probabilistic approach. Rips (2002) therefore suggests that reasoning theories should stick to their own proper domains.

4.2.2 Mental models

The mental models explanation of the suppression effects depends on the availability of counterexamples. Just as for the mental logic approach, mental models theory (Byrne *et al.*, 1999) suggests that people represent information about additional antecedents using *and*, and they represent information about alternative antecedents using *or*. This yields the following mental models representations for (6′): *if you turn the key* (*p*) **and** *there is fuel in the tank* (*q*), *then the car starts* (*r*) (additional) and for (7′): *if you turn the key* (*p*) **or** *it is hot-wired* (*q*), *then the car starts* (*r*) (alternative) (for the moment, ignore the models after the ellipsis, i.e. . . .):

And	*p*	*q*	*r*	*Or*	*p*		*r*
	. . .					*q*	*r*
						. . .	
	p	¬*q*	¬*r*		¬*p*	*q*	*r*

For (6′) a fully fleshed-out version of the mental model for *and* will include the case where the key is turned (*p* is true) but the car does not start (*not r* is true) because there is no fuel (*not q* is true). For (7′) a fully fleshed-out version of the mental model for *or* will include the case where the key is not turned (*not p* is true) but the car does start (*r* is true) because it was hot-wired (*q* is true). These models are shown after the ellipsis. This particular example shows why these counterexamples should be available. We have deliberately picked an example that appeals to your prior knowledge of cars and the factors that determine whether they start or not. This is called the *principle of pragmatic modulation* (Johnson-Laird and Byrne, 2002). That is, general knowledge in long-term memory can modulate the interpretation of conditionals, in this case making certain counterexamples much easier to access and represent.

Mental models also suggests that suppression effects should be all or nothing. However, some researchers working within the mental models framework (Quinn and Markovits, 2002; Schroyens and

Schaeken, 2003) have suggested that mental models should be supplemented with a *validating search procedure*. A conclusion is suggested by the mental model, and then long-term memory for world knowledge is searched to see if there is a counterexample. These can influence reasoning either in an all-or-nothing way (Quinn and Markovits, 2002) or in a graded, probabilistic way (Schroyens and Schaeken, 2003). Going down this route means that people can no longer be thought of as performing logical inferences. This is because if they interpret a conditional logically and assume it is true, then there can't be any counterexamples, i.e. *p* and *not-q* possibilities, to search for.

4.2.3 Probabilistic approach

Suppression effects in conditional inference are explained in terms of the effects of additional and alternative antecedents on the appropriate conditional probabilities (Oaksford and Chater, in 2010b). Suppose you are provided with an additional antecedent, or you retrieve one from memory. For example, suppose you are presented with or retrieve from memory the possibility that the petrol tank is empty. Your estimate of the probability that the car will start must be based on the proportion of times it has started given you have turned the key. Because an empty petrol tank will prevent your car from starting, then you must believe that this probability is smaller than your first estimate when all you knew was that the key was turned. That is, your estimate of the probability that the car starts is suppressed. In the suppression experiments people are not provided with the information in this way. Rather, they are given reminders about general preventive factors, like empty fuel tanks. According to the probabilistic approach, this has the effect of reducing people's estimates of the conditional probability of the car starting given you turn the key, $P(car\ starts|key\ turned)$. Thus, information about additional antecedents suppresses MP and MT inferences.

Explaining suppression effects for alternative antecedents follows a similar pattern. Alternative antecedents emphasize that, for example, it is possible to start cars without turning the key, e.g. by hot-wiring. They therefore suggest that the probability of the car starting given you don't turn the key, $P(s|\neg k)$, is higher than you first thought. This has to reduce the probability of the car not starting given that you do not turn the key, $P(\neg s|\neg k)$. This is simply because these probabilities must sum to 1, i.e. $P(s|\neg k) + P(\neg s|\neg k) = 1$. So if one goes up, the other must come down. The probability of the car not starting given that you do not turn the key is the probability that you must assess to determine whether to draw the DA inference. This is why alternative antecedents suppress the DA and AC inferences according to the probabilistic approach.

SUMMARY OF SECTION 4

In this section we have looked at two major experimental paradigms in conditional inference:

- the standard abstract task
- the suppression experiments.

How the various theories of the reasoning address these results was discussed:

- The abstract conditional inference task yields results that do not look consistent with logic.
- The *mental logic* theory explains these data by people using the more complex RAA inference to perform MT and by some people adopting a bi-conditional interpretation.
- The *mental model* theory also appeals to the bi-conditional interpretation but proposes initial semantic representations that can be 'fleshed out'.
- The *probabilistic approach* suggests that the data can be explained by people assessing the appropriate conditional probabilities.

For the suppression experiments:

- *Mental logic* could represent additional and alternative antecedents as conjunctive (and) or disjunctive (or) antecedents.
- *Mental models* proposes a similar solution and suggests that the principle of pragmatic modulation may make counterexamples more available.
- *The probabilistic approach* explains these effects by the way additional and alternative antecedents modulate the relevant conditional probabilities, which is compatible with the graded effects observed in this task.

The observation of graded effects has led to two responses from the other theories:

- *Mental logicians* argue for a clear distinction between deductive and non-deductive inference.
- Some *mental models* theorists attempt to incorporate these effects by introducing a probabilistic component.

This is one of the most active areas of current research in human reasoning.

5 WASON'S SELECTION TASK

The selection task is probably the most used task in the psychology of reasoning (Wason, 1968). We look first at the original *abstract* form of the task. As for the conditional inference task, we then look at how each theory explains the data.

5.1 The abstract selection task

In this version of the task people assess whether evidence is relevant to the truth or falsity of a conditional rule (Wason, 1968). In the abstract version, the rule concerns cards that have a number on one side and a letter on the other (Figure 12.2). A typical rule is *if there is an A on one side* (p), *then there is a 2 on the other side* (q). Four cards are placed before the participant, so that just one side is visible, showing an 'A' (p card), a 'K' (*not-p* card), a '2' (q card) and a '7' (*not-q* card) (Figure 12.2). Participants select those cards they must turn over to determine whether the rule is true or false. It was shown in Table 12.1 that sentences like *if p then q* are only false when the antecedent, p, is true (there is an A on one side) and the consequent, q, is false (there is not a 2 on the other side). Consequently, according to logic, only a card with an A on one side but without a 2 on the other side makes this rule false. There are only two cards that could possibly be of this type: the A card could have a number other than 2 on the other side, and the 7 card could have an A on the other side. So, according to logic people should select the A and the 7 cards to turn over, but *not* the K or the 2 cards.

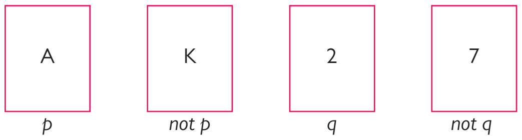


FIGURE 12.2 The four cards in the abstract version of the Wason selection task. Participants are told that each card has a letter on one side and a number on the other side. They are then asked to pick those cards that they must turn over to test whether the rule *if there is an A on one side then there is a 2 on the other side* is true or false.

These selections are rarely observed in the experimental results. Typical results are: *p* and *q* cards (46 per cent); *p* card only (33 per cent); *p*, *q*, and *not-q* cards (7 per cent); *p* and *not-q* cards (4 per cent); other card combinations (10 per cent) (Johnson-Laird and Wason, 1970). That is, as few as 4 per cent of participants select the response predicted by logic. Participants typically select cards that could *confirm* the rule, i.e. the *p* and *q* cards. However, according to logic, the choice of the *q* card is irrational, and is an example of so-called 'confirmation bias'. That is, people appear to be trying to find cards that have an A (*p*) on one side and a 2 (*q*) on the other side. Consequently, if we judge whether people are rational by a logical standard then these results seem to indicate that people are irrational.

The degree of people's irrationality was further illustrated by work on the *matching effect*. Evans and Lynch (1973) used rules that also contained negations, e.g. *if there is an A on one side then there is* ***not*** *a 2 on the other side*. Participants continued to select the *A* and the *2* card. For this rule these are now the logical responses because a card that does *not* have *not-2* on it must have a *2* on it. However, if they were showing confirmation bias they should now select the *A* and the *7* card. Selecting the *A* and the *2* card for both the standard and the negated rules seems only consistent with *matching bias*. That is, participants are not engaging rationally in the task at all but are simply matching the cards to the letters and numbers mentioned in the rule.

5.1.1 Mental logic

The mental logic approach to the selection task is identical to the conditional inference task. The card sides that participants can see act as the categorical premises in conditional inferences. So, for example, given the rule *if A then 2*, deciding whether to turn the *A* card is equivalent to drawing an MP inference to predict that there is a *2* on the other side. Moreover, suppose you adopt the bi-conditional interpretation, so you also believe that *if 2 then A*. Given this interpretation, deciding whether to turn the *not A* card is like drawing an MT inference to predict that there is a *not 2* on the other side (which is DA on the original rule). This account relies on the following identities: MP = *p* card, DA = *not-p* card, AC = *q* card, and MT = *not-q* card.

However, this account predicts that the proportion of people selecting the cards in the selection task should mirror the proportion of participants endorsing the corresponding conditional inferences. But this is not observed. In particular, while the *q* card is endorsed more than the *not-q* card in the selection task, the MT inference is endorsed more than the AC inference in the conditional inference task. However, one reason for this may be the way the categorical premises are presented in the selection task. Take MT. In the conditional inference task, people see *if A then 2*, *not-2*, therefore *not-A*. However, in the selection task, they see *if A then 2*, *7*, with no explicit conclusion. The point is that they have to infer that *7* is an instance of the category of numbers that are *not-2*. Presenting a negated premise in this way has been labelled an *implicit* negation (Evans *et al.*, 1996) and this is always the method used in the selection task. So what happens when implicit negations are used in the conditional inference task? The typical pattern for the *if A then 2* rule was MP = 95 per cent, DA = 38 per cent, AC = 79 per cent, and MT = 58 per cent (Evans and Handley, 1999), i.e. very close to the corresponding pattern of card selections in the selection task, given the identities above. This account may also explain the matching effect. If, for example, *not-2* is shown on the cards, rather than *7*, then the matching effect goes away (Evans *et al.*, 1996). So the matching effect may be a result of processing implicit negations and not a result of any inherent illogicality. In sum, it would appear that people may indeed be drawing conditional inferences in the selection task using the upturned face as the categorical premise.

5.1.2 Mental models

The explanation of the selection task results is very similar to the conditional inference task. However, now people consider each possibility for whether there could be something on the other side of the card that bears on the truth or falsity of the rule. We examine the card that should be turned for each interpretation in Table 12.2 (page 374).

✱ ACTIVITY 12.4

See if you can work out which cards should be selected for the bi-conditional interpretations and why (Table 12.2(c) and (d)). A clue is given by the identities between card selections and conditional inferences in the mental logic section.

If the conditional interpretation is adopted but it is not fleshed out ((a) in Table 12.2) then people will only turn the *A* (*p*) card. This is because if it has a *2* on the other side it confirms the rule but if it has a *7* (*not-q*) on the other side it falsifies the rule. The *2* card is not turned because whatever is on the other side, this card makes the rule true (refer back to the truth tables in Table 12.1). If the conditional interpretation is adopted and fleshed out ((b) in Table 12.2) then people will turn the *A* (*p*) and the *7* (*not-q*) cards. The reason for the selection of the *A* (*p*) card is the same as in (a). The *7* (*not-q*) card is now selected because it is represented and if it has an *A* (*p*) on the other side it falsifies the rule. The frequencies of card selections depend on the proportions of participants adopting the different interpretations.

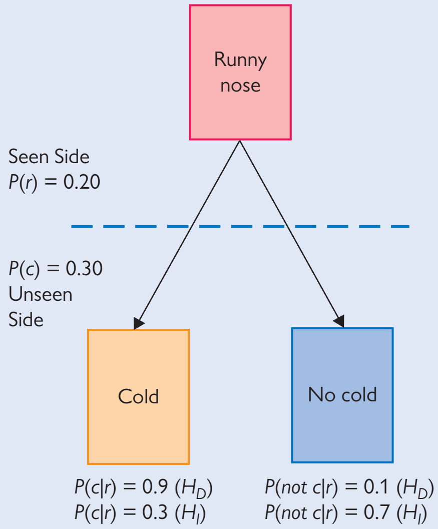
The matching effect is given exactly the same explanation in mental models theory as in mental logic. That is, it is a product of having to process implicit negations (Evans and Handley, 1999; Johnson-Laird and Byrne, 2002).

5.1.3 Probabilistic approach

The probabilistic approach suggests that selecting the *p* and the *q* card, far from being irrational, is the optimal response. The general idea is quite simple, although the mathematics can be a bit off-putting (Oaksford and Chater, 1994, 1996). We provide a worked example in Box 12.1. The idea is that in the selection task people are looking for the most informative evidence to decide whether John's having a runny nose is more often associated with having a cold than you would expect by chance. If it is, then there is a relationship between his runny nose and his having a cold so that his runny nose allows you to predict that he has a cold: they are *dependent*. If there is no such relationship, then they are *independent*. To determine which cards are more informative, information is quantified as *bits* of information. People are initially assumed to be maximally uncertain about the relationship between runny noses and colds. That is, the hypotheses that they are dependent and that they are independent are given an even chance (0.5) of being true. This means that people's uncertainty is the highest it can be at 1 bit. Turning cards to reveal data reduces this uncertainty. So, for example, turning the *runny nose* (*p*) card to find that *John did not have a cold* (*not-q*) on this occasion should reduce my uncertainty about which hypothesis is true. I should now be more certain that, for John, runny noses and colds are independent (see Box 12.1). This reduction in my uncertainty is called *information gain*. However, participants don't actually turn the cards over. So what gets calculated is *expected information gain*. This is the reduction in uncertainty averaged over the two possibilities, i.e. the other side of the *runny nose* card could reveal that John had a cold or that he did not, on this occasion.

Expected information gain can be calculated for each card (Box 12.1 shows the calculation for just the *p* card [*runny nose*]). Calculating the relevant probabilities of what is on the other side for each card involves the same three probabilities as in the conditional inference task. That is, the probability that John has a runny nose, $P(r)$, the probability that he has a cold, $P(c)$, and the probability that he has a cold given that he has a runny nose, $P(c|r)$. If $P(r)$ and $P(c)$ are both low (as in Box 12.1, i.e. 0.2 and 0.3 respectively) then the expected information gain for the *q* card (*John has a cold*) is higher than for the *not-q* card (*John does not have a cold*). Therefore, this model explains the standard finding as a *rational* consequence of seeking the most informative data. We can perhaps see why this happens intuitively using an alternative example rule: *if a pan drops in the kitchen, then it makes a clanging noise.* According to logic, testing this

BOX 12.1 Calculating expected information gain



H_D Dependence Hypothesis: if runny nose then cold $P(c|r) = 0.9$

H_I Independence Hypothesis: No relationship $P(c|r) = P(c) = 0.3$

Assume *runny nose* card is turned to reveal *no cold*; this is the data, *D*.

Belief before turning the card:
$P(H_D) = P(H_I) = 0.5$

Uncertainty Before: $I(H_i) =$

$$\sum_i P\,(Hi)\log_2\left(\frac{1}{P\,(H_i)}\right) = 1 \text{ bit}$$

Uncertainty After: $I(H_i|D) =$

$$\sum_i P(H_i|D)\log_2\left(\frac{1}{P(H_i|D)}\right)$$

$$P(H_D|D) = \frac{P(D|H_D)P(H_D)}{\sum_i P(D|H_i)P(H_i)} = \frac{0.1 \times 0.5}{0.1 \times 0.5 + 0.7 \times 0.5} = 0.125, \text{ and so } P(H_I|D) = 0.875,$$

i.e. highly probable H_I is true.

Therefore, uncertainty after $= 0.125 \log_2 (1/0.125) + 0.875 \log_2 (1/0.875) = 0.544$ bits

Amount by which uncertainty has reduced, or *Information Gain* $= I(H_i) - I(H_D|D) = 1 - 0.544 = 0.456$ bits. However, the card is never turned. So the *expected* uncertainty after turning the card is calculated. By similar means we can calculate the uncertainty after finding *cold* = 0.811. Expected uncertainty $= 0.5 \times (0.9 + 0.3) \times 0.811 + 0.5 \times (0.1 + 0.7) \times 0.544 = 0.704$.
Therefore, the expected reduction in uncertainty, or *expected information gain*, after turning the *runny nose* card $= 1 - 0.704 = 0.296$ bits.

hypothesis would involve investigating every instance of not hearing a clanging noise, to see whether a pan has dropped noiselessly so falsifying this rule. This clearly makes no sense *because* not hearing a clanging noise is a very rare event.

This account also explains the matching effect as a rational consequence of manipulating probabilities (Oaksford and Chater, 1994). The probability that John does not have a cold is greater than the probability that he does: most people are cold free most of the time. So, apart from being implausible, a rule like *if John has a runny nose, he does not have a cold*, introduces a high probability event in the consequent. It turns out that if either the antecedent or consequent has a high probability, then the expected information gain for the *not-q* card is higher than for the *q* card. So people should select the *John has a cold* card for this rule as well, which is the matching effect.

5.2 The deontic selection task

In this section we look at the task that has recently engendered the most controversy in reasoning research. This is largely because of the strong claim made by *evolutionary psychology* that this research reveals the effects of various innate cognitive modules. In this section we will therefore also consider this approach to human reasoning.

Some early work on the selection task seemed to show that logic-like performance was observed when *real-world* materials, as opposed to abstract materials, were used (Wason and Shapiro, 1971). People seemed to make more logical *p* and *not-q* card selections for rules like:

(8) *If Johnny travels to Manchester, he takes the train and*

(9) *If you use a 2nd class stamp, you must leave the envelope unsealed.*

However, rules like (8) only sometimes led to logical responses whereas rules like (9) and (10) reliably produced logic-like performance (Cheng and Holyoak, 1985).

(10) *If you are entering the country, you must have a cholera inoculation.*

For this rule, (10), participants were told to imagine that they were immigration officials at the airport and the cards represented immigration forms. The reason why rules like (9) and (10) reliably facilitate reasoning, but (8) does not, appears to have little to do with logic.

Rules like (9) and (10) are 'prescriptions' for how people should and should not behave; they are not descriptions of the world or of how someone behaves in it. This marks an important distinction (Manktelow and Over, 1987). Finding out that Johnny has travelled to Manchester by car may question the truth of (8). However, finding someone who is entering the country without a cholera inoculation does *not* question the *truth* of (10). (10) is in force regardless of the number of people found to be violating it. (8) is called an 'indicative' conditional, whereas (9) and (10) are 'deontic' conditionals: they describe rules for how people should or should not behave. Only these deontic conditionals appear to reliably produce logic-like performance. However, this is only because the task is no longer one for which standard logic, as we discussed above, provides the solution. That is, the reason why selecting the *p* and *not-q* cards is the correct action to perform has nothing to do with logic.

This point came sharply into focus when researchers began to investigate the factors that affect reasoning with deontic conditionals. Rule (10) is an *obligation* rule, i.e. it describes the pre-conditions (having an inoculation against cholera) that you are *obliged* to satisfy to carry out an action, i.e. entering the country. These deontic rules can also be framed as *permission* rules (Cosmides, 1989). So Rule (10) could be expressed as:

(11) *If you have a cholera inoculation, then you may enter the country.*

This rule now describes an action you are *permitted* to carry out given you have satisfied the pre-condition of having an inoculation against cholera.

There are two potentially important differences between (11) and (10). First, in (11) the pre-condition is now in the antecedent of the rule and the action is in the consequent. Second, in (11) *must* has changed to *may* in the consequent to reflect the fact that you are not obliged to enter the country if you have a cholera inoculation but you *may* if you want to. This first difference has an important consequence. If you are an immigration official, regardless of how the rule is expressed, i.e. (11) or (10), you are still concerned to find people who are breaking the immigration laws, i.e. people who are trying to enter the country without a cholera inoculation. For (10), as we have seen, this corresponds to the logical *p* and *not-q* cards. However, for (11) entering the country corresponds to the true consequent case, i.e. *q*, and not having a cholera inoculation corresponds to the false antecedent case, i.e. *not-p*. Therefore, if participants are looking for potential law breakers, as they are asked to do in this task, then they should select the *not-p* and *q* cards when given a rule like (10). Cosmides (1989) observed exactly this behaviour. This pattern of responses would appear to have nothing to do with the logic of the conditional. Other manipulations have revealed similar performance (Manktelow and Over, 1991).

In sum, the deontic selection task seemed to reveal the conditions under which people could reason logically. However, the factors that affect deontic reasoning, i.e. the nature of the rules, *obligation* vs *permission*, revealed that logical reasoning was not being facilitated. This is because the *not-p* and *q* card selection cannot be predicted from logic, although it makes perfect sense for the rules introduced in the deontic selection task. We now look at how these findings have been interpreted by the different psychological theories of reasoning.

5.2.1 Mental logic

Mental logicians have not explicitly addressed the deontic selection task. However, philosophers and mathematicians have formulated logical theories that deal with the 'modal' terms, e.g. *must* and *may*, which feature in deontic conditionals. So, mental logic might be extended to these inferences in the future (Manktelow and Over, 1987).

5.2.2 Mental models

In its basic principles the mental models account of the deontic selection task is continuous with the explanation provided for the standard selection task and indeed for all reasoning. People represent possibilities and identify counterexamples. The crucial distinction for explaining the deontic selection task is that what is deontically possible or permissible does not correspond to what is factually possible (Johnson-Laird and Byrne, 2002). So for example, if it is true that *if I turn the key then the car starts*, then the three models in Table 12.2(b) are factually possible, but the model in which the key is turned but the car does not start is not possible. However, take the rule (10) *If you are entering the country, you must have a cholera inoculation*. For this rule, all four truth table cases are factually possible – even the case where someone enters the country without a cholera inoculation. The deontic rule does not say that this *does* not happen, it says that it *should* not happen, i.e. it is factually possible but deontically impermissible. This means that the representation of (10) is as follows:

10

Factual possibility	*enter*	***inoculation***	*Deontic possibility*
Factual possibility	*¬enter*	***inoculation***	*Deontic possibility*
Factual possibility	*¬enter*	***¬inoculation***	*Deontic possibility*
Factual possibility	*enter*	***¬inoculation***	*Deontic impossibility*

That is, people explicitly label or mentally tag the different possibilities indicating whether they are permissible (deontically possible) or impermissible (deontically impossible). Exactly the same mental representation is formed of (11) *If you have a cholera inoculation, then you may enter the country.* So in both cases people are looking for cases that are deontically impossible. This explains why the logically different cases are selected for the rules (10) and (11). Even though different logical cases are being selected because of the action and precondition switching between antecedent and consequent, the same deontic case is being selected (Johnson-Laird and Byrne, 2002).

5.2.3 The probabilistic approach

The probabilistic approach to the deontic selection task adopts the decision theoretic approach first proposed by Manktelow and Over (1987, 1991). In decision theory (see Chapter 11) people make choices that help to maximize *expected utility*, where utilities are the values people place on various outcomes. In the deontic selection task, the instructions ask people to place a high value on instances of unfairness, e.g. where someone enters the country without a cholera inoculation.

Oaksford and Chater's (1994) formal model can be very easily illustrated by annotating probabilities and utilities in a mental model (Johnson-Laird *et al.*, 1999) of the possible outcomes in a deontic selection task:

		Utilities	*Probabilities*
enter	***inoculation***	*−0.1*	*0.25*
¬enter	***inoculation***	*−0.1*	*0.25*
¬enter	***¬inoculation***	*−0.1*	*0.25*
enter	***¬inoculation***	*5*	*0.25*

When you are told that you are an enforcer of the regulation that *if you are entering the country, then you must have a cholera inoculation*, you realize that your goal is to find people who are entering the country but who do not have an inoculation against cholera. You therefore assign a high positive utility of 5 to this case. All other possibilities are not interesting so they are assigned a small negative utility of −0.1. Assume that on the particular plane that has just landed you have no knowledge either of how likely people are to have a cholera inoculation or how likely they are to be trying to enter the country. Consequently the four possibilities are equally likely. You can now calculate the expected utility associated with turning each card. So for the card marked *enter* you need to consider the probability that this person has or has not got a cholera inoculation. Therefore you need to calculate two probabilities. The probability of the person having an inoculation given they are trying to enter is 0.5, i.e. the probability of entering with an inoculation (0.25) divided by the probability of having a cholera inoculation (0.5). A similar calculation gives the same values for the probability that the person does *not* have an inoculation given that they are trying to enter. These probabilities are then multiplied by the corresponding utilities and summed to provide the expected utility (EU) associated with turning the card:

$$\begin{aligned} EU(enter) &= P(inoculation \mid enter) \times U(enter, \\ &\quad inoculation) + P(\neg inoculation \mid enter) \\ &\quad \times U(enter, \neg inoculation) \\ EU(enter) &= (0.5 \times 5) + (0.5 \times -0.1) = 2.45 \end{aligned}$$

Similar calculations can be carried out for each card:

$$\begin{aligned} &EU(\neg enter) = (0.5 \times -0.1) + (0.5 \times -0.1) = -0.1 \\ &EU(inoculation) = (0.5 \times -0.1) + (0.5 \times -0.1) = -0.1 \\ &EU(\neg inoculation) = (0.5 \times 5) + (0.5 \times -0.1) = 2.45 \end{aligned}$$

According to the principle of maximizing expected utility, you should therefore pick the *enter* (*p*) card and the *not-inoculation* (*not q*) card.

This decision theoretic account gives the same answers as mental models theory. However, it suggests that people's deontic reasoning should be sensitive to more than the deontic permissibility of the available possibilities. It suggests that manipulations of utility and probability should also affect people's deontic reasoning in a graded manner. There is evidence that seems to support these predictions (Kirby, 1994; Manktelow *et al*., 1995).

5.2.4 Evolutionary psychology

Evolutionary psychology is a general movement in psychology that sees many of our cognitive mechanisms as innately specified, having evolved to cope with adaptive problems confronted by early humans. Evolutionary psychologists have argued that various aspects of human reasoning might also be under the control of innately specified cognitive modules (Cosmides, 1989; Fiddick *et al*., 2000). In particular, they suggest that deontic reasoning is one such domain.

Many of the effects observed in the deontic selection task can be explained by cognitive modules for *social contracts* modules (Cosmides, 1989; Fiddick *et al*., 2000). This domain was vital in the past when early humans were evolving in social groups. A social exchange situation involves satisfying some kind of requirement in order to receive a benefit from some other individual or group. This can be expressed as the social contract: *if you accept the benefit then you must satisfy the requirement*. An equivalent form of this rule is *if you satisfy the requirement then you are entitled to the benefit*. The cognitive module specialized for reasoning about social contracts is largely insensitive to the logic of the conditionals used to describe them. What is important for survival out on the Savannah is that you do not get *cheated* in these social exchange situations. Consequently, you should be on the lookout for people who take the benefit but who do not meet the requirement. Of course, for the *if you are entering the country, then you must have a cholera inoculation* rule, this corresponds to someone trying to enter the country without an inoculation against cholera. Consequently, this account can explain the results of the standard deontic selection task.

What distinguishes this account from other explanations? The point of invoking cognitive modules is that they are automatically invoked and will tend to override any domain-general reasoning abilities we may possess. Moreover, because they are domain specific, they should only apply to situations where there is a clear *benefit* and *requirement* involved. Cosmides (1989) constructed two task versions using rules like *if* *a student is to be assigned to Grover High School, then that student must live in Grover city*. In one version of the task participants were told that going to Grover High was a benefit compared to going to Hanover High (the *not-p* case). In another version this information was not included, so although one is obliged to live in Grover city to go to Grover High, there is no suggestion that going to Grover High is of any benefit to you. Far more people selected the *p* and *not-q* cards when the benefit was explicitly mentioned than when it was not. Consequently, it would appear that the obligation rule form is not sufficient to produce the *p* and *not-q* response; the *p* and the *q* must be understood as a benefit and as a requirement respectively.

Further experiments appeared to show that people have an automatic understanding of social exchange situations in the absence of any rules (Fiddick *et al.*, 2000). In one condition, participants were given the rule *if you give me some potatoes, then I will give you some corn*. In another condition participants were told to imagine they were a farmer who walks into the neighbouring village and meets someone who says *I want some potatoes*, to which you respond *I want some corn*. Participants are then given four cards corresponding to four people marked: *you gave this person potatoes*, *you gave this person nothing*, *this person gave you corn*, and *this person gave you nothing*. Participants are asked to check whether any of the people represented by the cards have cheated them. Both groups performed equally well. Now this could be because people interpret the rule-less version as involving the rule. However, Fiddick *et al.* (2000) observed that people could translate the rule-less scenario into any one of four different underlying rules, any one of which would be consistent with a social exchange, but only one of which could produce the observed results. Consequently, it would seem that reliable deontic selection task performance requires the appropriate *benefits* and *requirements* to be specified but is independent of the use of a conditional rule. The latter excludes any explanation of these tasks where the logic of the conditional plays a role. However, this interpretation has been the subject of intense debate (Sperber and Girotto, 2002).

SUMMARY OF SECTION 5

In this section we have looked at two major reasoning paradigms using the Wason selection task:

- the standard abstract task and the matching effect
- the deontic selection task.

How the various theories of reasoning address these results was discussed. For the abstract task:

- *Mental logic* suggests that people use the card face they can see to draw conditional inferences.
- *Mental models* suggests that people check their mental models for cases that might bear on the truth or falsity of the rule.
- *The probabilistic approach* suggests that people are seeking the most informative evidence to determine whether *p* and *q* are dependent or not.

For the deontic selection task:

- *Mental logic* has not been extended to these data, although there are logics for this mode of reasoning.
- *Mental models* proposes that people label the various factual possibilities with whether they are deontically permissible or impermissible.
- *The probabilistic approach* explains these effects by the way that the context affects the utilities associated with the various possible outcomes of turning a card.
- *Evolutionary psychology* proposes that we possess innate cognitive modules for reasoning about social contracts that are independent of the conditional rules.

The current controversy concerns whether domain-general accounts can explain these results or whether domain-specific mechanisms are needed.

6 DISCUSSION

We have seen how these different theories account for the most researched experimental tasks in human reasoning. The ability to explain these results is one main criterion by which to judge these theories. As we saw they all fared reasonably well. In this final section, we continue evaluating these different theories, in particular for what they have to say about the issue of human rationality. We also take the opportunity to introduce some further evidence that might help decide between these theories. In evaluating theories, there are two possible approaches, competitive and integrative. In a competitive approach, each theory is regarded as in competition to be the one true theory of reasoning. The idea is that the proponents of each theory fight their own corner, attempting to find the killer argument or experiment that will support their theory and falsify all the others. This approach tends to lead to acrimonious exchanges in the literature. However, rarely is any argument or evidence regarded as fatal. Indeed, in this and in many other areas of psychology, such wrangles usually end up in some kind of compromise position where it is conceded that each theory probably has its own merits and proper domain of application. Consequently, the final position arrived at is often an *integration* of theoretical positions. In this light, we first look at the relative merits of each theory before closing this chapter by looking to integrative approaches.

6.1 Theoretical evaluation

In this section we look at each theory and examine further evidence to distinguish between these theories where it exists. However, this evaluation will fall short of plumping for one theory over another.

6.1.1 Mental logic

Mental logic theories have several advantages:

- They are formally very well specified. This means that theoreticians are able to prove mathematically what can be predicted by these theories.
- They preserve a full logical conception of what it is to reason rationally.

However, they also have some disadvantages:

- It is unclear how mental logic can apply to a range of data on, for example, the graded phenomena observed in the suppression experiments.
- Consequently, the range of coverage of the theory is quite narrow.

Recently, this has led some mental logicians (Rips, 2001, 2002) to propose a sharp distinction between *non-deductive* and *deductive* reasoning. The former are evaluated in terms of how probable the premises make the conclusion, which Rips calls *inductive strength*. People may be able to evaluate arguments for both *inductive strength* and *deductive validity*. This last issue points to a possible integration whereby mental logic deals with clear-cut cases of deductive reasoning and theories like the probabilistic approach deal with the rest.

6.1.2 Mental models

Features in favour of mental models include:

- The range of coverage of the data. For many phenomena in human reasoning mental models provides the only existing account. This suggests that while mental logic may have the advantage of depth, mental models has the advantage of breadth.
- More reasoning researchers work in this framework than in any other.

However, as with mental logic there are some problems:

- Pragmatic modulation does not seem consistent with graded effects. If a counterexample is accessed then the inference should not be drawn, and if it is not accessed then it should be drawn.

There are also some more general issues:

- The attempt to provide the 'crucial experiment' that clearly falsifies mental logic while supporting mental models has largely been unsuccessful (although this is largely in the eye of the beholder and some recent work may be argued to play this role; Johnson-Laird and Savary, 1999).

It is difficult to gauge what mental models theory says about human rationality. Although initially motivated by the logical meanings associated with the structure-building words, mental models theory has been extended well beyond the scope of standard logic. Within the scope of standard logical inference, mental models can be seen as preserving human rationality because it *approximates* logical reasoning. However, beyond the scope of standard logical inference, mental models theorists rarely show that their theories approximate any logical or mathematical theory of reasoning. Consequently, in these domains it is difficult to tell whether the theory preserves human rationality or not. This is where the mental logic theory and the probabilistic approach agree – both attempt to preserve human rationality by showing that most reasoning behaviour approximates to either logic or probability theory.

The cognitive neuroscience of reasoning may also address the issue of whether people reason with a language-based mental logic or more imagery-based mental models. For example, there is a neuroimaging study that analysed people while they performed conditional inference and other reasoning tasks (Goel *et al.*, 1998). It was found that activation was primarily restricted to the left hemisphere language centres rather than the right hemisphere imagery systems. Results like this seem to argue for a mental logic approach. More recently, Kroger *et al.* (2008) have shown that regions in the right prefrontal cortex and inferior parietal lobe were more active in reasoning, consistent with an imagery-based mental models approach. However, such data is far from conclusive and there are serious difficulties in localizing reasoning ability in the brain (Goel, 2007). Nonetheless, this is certainly an interesting future direction for reasoning research.

6.1.3 Probabilistic approach

The general advantages of the probabilistic approach are:

- Much more of human reasoning behaviour can be seen as rational but account must be taken of people's prior knowledge of the environment, e.g. the rarity assumption.
- Predictions can be derived for how manipulating probabilities and utilities should affect reasoning performance. These have generally been confirmed.

The disadvantages of the probabilistic approach are:

- The coverage of the probabilistic approach is small compared to the mental models approach.
- The theory only provides an account of how the cognitive system should behave given certain inputs. It does not provide an account of the cognitive representations and processes involved that some feel is the proper level of psychological explanation.

It is important to bear in mind that this theory suggests that the standard of rationality should change. Rather than judge human reasoning by logical standards, it should be judged by a probabilistic standard. When it is, a lot more of people's behaviour can be viewed as rational than the early experiments on human reasoning led us to expect.

6.1.4 Evolutionary psychology

It is difficult to judge the evolutionary psychology approach by the same standards as the other theories because even compared to the mental logic and probabilistic approaches its scope is extremely limited. However, within the domain of deontic reasoning there is some further neuropsychological evidence that may be relevant. The evolutionary approach suggests that people have two innate cognitive modules, one for social contracts and one for reasoning about *hazard management*. The latter involves reasoning about rules like *if you clear up blood, you must wear rubber gloves* (Manktelow and Over, 1991). The rule indicates the precautions you should take given that you encounter a hazardous situation. According to the domain-general theories, such rules are dealt with by the same mechanisms that deal with social contract rules. However, there is recent evidence of a neuropsychological patient with brain damage who shows an impaired ability to reason about social contracts but an intact ability to reason about hazard management rules (Stone *et al.*, 2002). This seems to suggest that this patient has an intact innate hazard management module but a damaged innate social contract module. However, unless a patient is found with the opposite deficit, i.e. impaired hazard management reasoning and intact social contract reasoning, these results remain inconclusive.

6.2 Integrative approaches

One of the most important recent developments in the psychology of reasoning has been the attempt to see how far an integrative approach is possible that appeals to more than one theory. Indeed, in a recent edited collection of work on conditional inference (Oaksford and Chater, 2010a), nearly a third of the chapters took an explicitly integrative approach. Two pre-existing strands in the psychology of reasoning have underpinned these developments: *dual process theories* and work on *individual differences*.

6.2.1 Dual process theories

Dual process theories have a long pedigree in reasoning research (Evans, 1984, 2007; Evans and Over, 1996; Stanovich, 2011; Stanovich and West, 2000). These theories suggest a two-way partition in reasoning abilities. As we have already suggested, this is similar to the ideas of some mental logic theorists who have invoked the distinction between deductive and non-deductive reasoning (Rips, 2002). Typically these theories suggest that we do have a, perhaps limited, ability for explicit logical reasoning that may be embodied in a mental logic or in mental models. However, a lot of reasoning goes on implicitly and is independent of these logical processes. More recently the distinction has been drawn between two types of rationality (Evans and Over, 1996). People are rational in one sense when their reasoning conforms to a normative standard like logic. They are rational in another sense when they reason so as to achieve their goals in the world, regardless of whether their reasoning conforms to a normative standard. Different mental processes are involved in these forms of reasoning.

One way of conceiving of the distinction between the processes involved is in terms of what Evans (2007) refers to as Type 1 and Type 2 processing. Type 1 consists of long-term memory (LTM) processes (see Chapter 8) involving associations and heuristics. Type 2 consists of analytic processes that are carried out in working memory (WM), which is a late evolutionary development (see Chapter 9). That is, these processes go on in different parts of the *cognitive architecture* (see Chapter 19). (NB This is wholly different to the use of the terms Type I and Type II processing in relation to long-term memory – see Chapter 8.)

6.2.2 Individual differences

An important recent source of evidence for a dual process approach is the study of individual differences (Stanovich, 2011; Stanovich and West, 2000). For example, it has been shown that the ability to make the logically correct response on the selection task is associated with IQ (Stanovich and West, 1998). It would appear that participants with high IQ are capable of interpreting this task logically. However, if these participants are removed then IQ seems to correlate with the standard *p* and *q* card response (Newstead *et al.*, 2004). This evidence seems to argue for a dual process theory. People possess automatic unconscious reasoning mechanisms that operate in accordance with probabilistic standards of reasoning. However, those with higher IQs are capable of ignoring the prior knowledge that is required to determine the relevant probabilities, and of reasoning logically about the task.

6.2.3 Direct evidence for integrative approaches

Other than the relationship with individual difference factors like IQ, there is increasing experimental evidence that within individuals both processes may be observed. We mention two sources of such evidence for dual processing. First, there is evidence that Type 1 processes involving prior probabilistic beliefs held in LTM are produced rapidly, but when people take more time they produce responses more consistent with analytic Type 2 processes (Verschueren *et al.*, 2005; but see Handley *et al.*, 2011). Second, in an ingenious neuroimaging experiment, De Neys *et al.* (2008) monitored the brain regions believed to be involved in conflict detection (anterior cingulate) when people either produced a Type 1 belief-biased response not consistent with the Type 2 analytic response, or when they produced a Type 2 analytic response. They found that conflict detection centres were active in the brain when both responses were produced. So, even when responding reflexively via Type 1 processes, people seem unconsciously to 'know' that there is conflict with the response they 'should' produce via Type 2 analytic processes.

6.2.4 Modelling integrative approaches

Other evidence for the particular processes involved comes from attempts to model the underlying cognitive processes (see Chapter 19) to see which theory or combination of theories provides the best fit to the data. For example, Oberauer (2006) tested a variety of theories by combining them using something called a *multinomial processing tree model*. The fine detail need not concern us. What this approach does is to assign values to how likely participants are to make particular processing choices and combinations of choices. The model homes in on those values that will result in the observed patterns of data. This approach also provides an index of how well each model fits the data, i.e. how well each model predicts each data point. The conclusion was that a particular dual process account provided better fits to the data than, in particular, the probabilistic approach. However, this may be an odd comparison to make, as for many proponents of dual process theories the probabilistic approach provides the theory of Type 1 processes. More recent modelling of dual processes (Oaksford and Chater, 2010b) shows how given the WM/LTM cognitive architecture assumed in almost all cognitive theory, implementing the probabilistic approach will require WM representations like those proposed by mental models theory (see also Evans, 2007).

In conclusion, while the jury is still very far from delivering a verdict on the correct theory of human reasoning, the recent emergence of integrative approaches seems like a very positive development. This marks out the last ten years of research as something of a paradigm shift towards what Over (2009) refers to as a *new paradigm psychology of reasoning*.

FURTHER READING

Evans, J.St.B.T. (2007) *Hypothetical thinking: Dual Processes in Reasoning and Judgment*, Hove, UK, Psychology Press. The definitive statement of dual process theory by its originator, Jonathan Evans.

Johnson-Laird, P.N. (2006) *How We Reason*, Oxford, Oxford University Press. The most recent book-length treatment of Johnson-Laird's long-standing mental models theory.

Oaksford, M. and Chater, N. (2007) *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*, Oxford, Oxford University Press. The definitive statement of Oaksford and Chater's probabilistic approach, which provides the philosophical motivation and origins in the instructive failures of artificial intelligence.

Stanovich, K.E. (2011) *Rationality and the Reflective Mind*, Oxford, Oxford University Press. The most recent summary of Keith Stanovich's dual process approach, richly integrated into contemporary philosophy of rationality and mind.

REFERENCES

Boole, G. (1854) *An Investigation of the Laws of Thought on which are Founded the Mathematical Theories of Logic and Probabilities*, Cambridge, UK, Macmillan & Co.

Braine, M.D.S. and O'Brien, D.P. (eds.) (1998) *Mental Logic*, Mahwah, NJ, Lawrence Erlbaum Associates.

Byrne, R.M.J. (1989) 'Suppressing valid inferences with conditionals', *Cognition*, vol.31, pp.1–21.

Byrne, R.M.J., Espino, O., and Santamaria, C. (1999) 'Counterexamples and the suppression of inferences', *Journal of Memory and Language*, vol.40, pp.347–73.

Cheng, P.W. and Holyoak, K.J. (1985) 'Pragmatic reasoning schemas', *Cognitive Psychology*, vol.17, pp.391–416.

Cosmides, L. (1989) 'The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task', *Cognition*, vol.31, pp.187–276.

Cummins, D.D. (1995) 'Naive theories and causal deduction', *Memory and Cognition*, vol.23, no.5, pp.646–58.

Cummins, D.D., Lubart, T., Alksnis, O., and Rist, R. (1991) 'Conditional reasoning and causation', *Memory & Cognition*, vol.19, pp.274–82.

De Neys, W., Vartanian, O., and Goel, V. (2008) 'Smarter than we think: when our brains detect that we are biased', *Psychological Science*, vol.19, pp.483–9.

Ebbinghaus, H. (1885) *Memory* (H.R. and C.E. Bussenius, trans.), New York, Teachers College, Columbia University (1913).

Evans, J.St.B.T. (1977) 'Linguistic factors in reasoning', *Quarterly Journal of Experimental Psychology*, vol.29, pp.297–306.

Evans, J.St.B.T. (1984) 'Heuristic and analytic processes in reasoning', *British Journal of Psychology*, vol.75, pp.451–68.

Evans, J.St.B.T. (2007) *Hypothetical Thinking: Dual Processes in Reasoning and Judgment*, Hove, UK, Psychology Press.

Evans, J.St.B.T. and Handley, S.J. (1999) 'The role of negation in conditional inference', *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, vol.52A, no.3, pp.739–69.

Evans, J.St.B.T. and Lynch, J.S. (1973) 'Matching bias in the selection task', *British Journal of Psychology*, vol.64, pp.391–7.

Evans, J.St.B.T. and Over, D.E. (1996) *Rationality and Reasoning*, Hove, UK, Psychology Press.

Evans, J.St.B.T., Clibbens, J., and Rood, B. (1996) 'The role of implicit and explicit negation in conditional reasoning bias', *Journal of Memory and Language*, vol.35, no.3, pp.392–409.

Fiddick, L., Cosmides, L., and Tooby, J. (2000) 'No interpretation without representation: the role of domain-specific representations and inferences in the Wason selection task', *Cognition*, vol.77, no.1, pp.1–79.

Goel, V. (2007) 'The anatomy of deductive reasoning', *Trends in Cognitive Sciences*, vol.11, pp.435–41.

Goel, V., Gold, B., Kapur, S., and Houle, S. (1998) 'Neuroanatomical correlates of human reasoning', *Neuropsychologia*, vol.29, pp.901–9.

Handley, S.J., Newstead, S.E., and Trippas, D. (2011) 'Logic, beliefs, and instruction: a test of the default interventionist account of belief bias', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.37, pp.28–43.

Inhelder, B. and Piaget, J. (1958) *The Growth of Logical Reasoning*, New York, Basic Books.

Johnson-Laird, P.N. (1983) *Mental Models*, Cambridge, Cambridge University Press.

Johnson-Laird, P.N. (2006) *How We Reason*, Oxford, Oxford University Press.

Johnson-Laird, P.N. and Byrne, R.M.J. (1991) *Deduction*, Hillsdale, NJ, Erlbaum.

Johnson-Laird, P.N. and Byrne, R.M.J. (2002) 'Conditionals: a theory of meaning, pragmatics, and inference', *Psychological Review*, vol.109, no.4, pp.646–78.

Johnson-Laird, P.N. and Savary, F. (1999) 'Illusory inferences: a novel class of erroneous deductions', *Cognition*, vol.71, no.3, pp.191–229.

Johnson-Laird, P.N. and Wason, P.C. (1970) 'Insight into a logical relation', *Quarterly Journal of Experimental Psychology*, vol.22, no.1, pp.49–61.

Johnson-Laird, P.N., Legrenzi, P., Girotto, V., Legrenzi, M.S., and Caverni, J.P. (1999) 'Naive probability: a mental model theory of extensional reasoning', *Psychological Review*, vol.106, no.1, pp.62–88.

Kirby, K.N. (1994) 'Probabilities and utilities of fictional outcomes in Wason's four card selection task', *Cognition*, vol.51, pp.1–28.

Kroger, J.K., Nystrom, L.E., Cohen, J.D., and Johnson-Laird, P.N. (2008) 'Distinct neural substrates for deductive and mathematical processing', *Brain Research*, vol.1243, pp.86–103.

Manktelow, K.I. and Over, D.E. (1987) 'Reasoning and rationality', *Mind and Language*, vol.2, pp.199–219.

Manktelow, K.I. and Over, D.E. (1991) 'Social roles and utilities in reasoning with deontic conditionals', *Cognition*, vol.39, pp.85–105.

Manktelow, K.I., Sutherland, E.J., and Over, D.E. (1995) 'Probabilistic factors in deontic reasoning', *Thinking and Reasoning*, vol.1, pp.201–20.

Newstead, S.E., Handley, S.J., Harley, C., Wright, H., and Farrelly, D. (2004) 'Individual difference in deductive reasoning', *Quarterly Journal of Experimental Psychology*, vol.57, pp.33–60.

Oaksford, M. and Chater, N. (1994) 'A rational analysis of the selection task as optimal data selection', *Psychological Review*, vol.101, pp.608–31.

Oaksford, M. and Chater, N. (1996) 'Rational explanation of the selection task', *Psychological Review*, vol.103, pp.381–91.

Oaksford, M. and Chater, N. (1998) *Rationality in an Uncertain World: Essays on the Cognitive Science of Human Reasoning*, Hove, Psychology Press.

Oaksford, M. and Chater, N. (2003a) 'Computational levels and conditional inference: reply to Schroyens and Schaeken (2003)', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.29, no.1, pp.150–6.

Oaksford, M. and Chater, N. (2003b) 'Optimal data selection: revision, review and re-evaluation', *Psychonomic Bulletin and Review*, vol.10, pp.289–318.

Oaksford, M. and Chater, N. (2007) *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*, Oxford, Oxford University Press.

Oaksford, M. and Chater, N. (2009) 'Precis of Bayesian rationality: the probabilistic approach to human reasoning', *Behavioral and Brain Sciences*, vol.32, pp.69–120.

Oaksford, M. and Chater, N. (eds.) (2010a) *Cognition and Conditionals*: *Probability and Logic in Human Thinking*, Oxford, Oxford University Press.

Oaksford, M. and Chater, N. (2010b) 'Conditionals and constraint satisfaction: reconciling mental models and the probabilistic approach?' in Oaksford, M. and Chater, N. (eds.) *Cognition and Conditionals*: *Probability and Logic in Human Thinking*, Oxford, Oxford University Press.

Oaksford, M., Chater, N., and Larkin, J. (2000) 'Probabilities and polarity biases in conditional inference', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.26, pp.883–99.

Oberauer, K. (2006) 'Reasoning with conditionals: a test of formal models of four theories', *Cognitive Psychology*, vol.53, pp.238–83.

Over, D.E. (2009) 'New paradigm psychology of reasoning', *Thinking & Reasoning*, vol.15, pp.431–8.

Quinn, S. and Markovits, H. (2002) 'Conditional reasoning with causal premises: evidence for a retrieval model', *Thinking and Reasoning*, vol.8, no.3, pp.179–91.

Rips, L.J. (1994) *The Psychology of Proof*, Cambridge, MA, MIT Press.

Rips, L.J. (2001) 'Two kinds of reasoning', *Psychological Science*, vol.121, no.2, pp.129–34.

Rips, L.J. (2002) 'Reasoning imperialism', in Elio, R. (ed.) *Common Sense, Reasoning and Rationality*, New York, Oxford University Press.

Schroyens, W. and Schaeken, W. (2003) 'A critique of Oaksford, Chater, and Larkin's (2000) conditional probability model of conditional reasoning', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.29, no.1, pp.140–9.

Sperber, D. and Girotto, V. (2002) 'Use or misuse of the selection task? Rejoinder to Fiddick, Cosmides, and Tooby', *Cognition*, vol.85, no.3, pp.277–90.

Stanovich, K.E. (2011) *Rationality and the Reflective Mind*, Oxford, Oxford University Press.

Stanovich, K.E. and West, R.F. (1998) 'Cognitive ability and variation in selection task performance', *Thinking and Reasoning*, vol.4, no.3, pp.193–230.

Stanovich, K.E. and West, R.F. (2000) 'Individual differences in reasoning: implications for the rationality debate?' *Behavioral and Brain Sciences*, vol.23, no.5, pp.645–64.

Stone, V.E., Cosmides, L., Tooby, J., Kroll, N., and Knight, R.T. (2002) 'Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage', *Proceedings of the National Academy of Sciences*, vol.99, no.17, pp.11 531–11 536.

Taplin, J.E. (1971) 'Reasoning with conditional sentences', *Journal of Verbal Learning and Verbal Behaviour*, vol.10, pp.219–25.

Verschueren, N., Schaeken, W., and d'Ydewalle, G. (2005) 'A dual-process specification of causal conditional reasoning', *Thinking & Reasoning*, vol.11, pp.239–78.

Wason, P.C. (1968) 'Reasoning about a rule', *Quarterly Journal of Experimental Psychology*, vol.20, pp.273–81.

Wason, P.C. and Shapiro, D. (1971) 'Natural and contrived experience in a reasoning problem', *Quarterly Journal of Experimental Psychology*, vol.23, pp.63–71.

@ Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

PART 5

CONTRIBUTIONS FROM THE COGNITIVE NEUROSCIENCES

Introduction

13 Cognitive neuropsychology
Gianna Cocchini

14 Neuroimaging
Ingrid Johnsrude and Olaf Hauk

15 Cognitive neuropsychiatry
David Okai, Angus Gellatly, and Anthony David

INTRODUCTION

This part of the book represents a shift in emphasis. While the preceding chapters have each focused on a topic within cognition and explored what the research of cognitive psychologists has revealed, the next three chapters all, in one way or another, detail a different approach to investigating cognition. It is an approach that brings together a focus on cognition with research programmes that are typically described as instances of neuroscience.

Why should a cognitive psychologist be interested in neuroscience? A simple answer is one that we have already briefly touched on in Chapter 1. If cognition is a form of complex information processing then we can understand it at any one of Marr's (1982) three levels of analysis. At level 1, the computational level, cognition is understood in terms of what the mind computes and why. Visual perception, for example, might compute a function mapping patterns of light that fall onto the retina (the inputs to the function) into a description of the scene this corresponds to, the objects within the scene, their spatial relations, and so on (the outputs of the function). Why should cognition compute such a function? Because doing so recovers true (and valuable) information about what things lie in our immediate visually perceptible environment. At level 2, the algorithmic level, the cognitive psychologist is concerned with specifying the processing steps that intervene between these inputs and outputs, trying to understand how it is that the mind can achieve this recovery of useful information. At the hardware or implementational level, level 3, the focus is on how the processing steps specified at level 2 are physically realized in the human brain and body.

Some scientists and philosophers would argue that this last question is of little direct concern to cognitive psychology, and you can read some of these arguments later on in Chapter 20. However, these arguments can easily be challenged. Marr himself believed that the hardware level, level 3, constrains the kinds of algorithmic processing steps that may be found at level 2. Put another way, suppose that we found out that a particular model of visual perception could not be implemented in the brain. No matter how well the model might have been developed and argued, learning that it could not possibly be implemented in a human brain would be *prima facie* evidence that human beings don't achieve visual perception in the way the model proposes. The constraint from the hardware level can also work the other way around. It is possible that we discover interesting properties of the brain that lead us to believe some classes of algorithm are more plausible than others. Connectionist models are predicated on this idea – since they are built on an–admittedly–simple model of how the brain processes information, to the extent that models of cognition can be implemented as connectionist networks, such models are thought to be more plausible candidates for how the mind actually works.

Though this discussion is abstract, in this part of the book it becomes much more concrete. Each of the three chapters shows how paying attention to work within neuroscience contributes greatly to our understanding of cognition and cognitive processes. However, there is a further shift in emphasis that should be noted. Thus far, we have not explicitly focused on the methods that cognitive psychologists use. You may be much relieved that we have not – after all, learning about experimental method, research design, statistical analysis, and the like can seem dry and dusty. We have tended to assume that in studying cognitive psychology you will already have encountered such methodological rudiments in other accompanying courses, and so we have tended to assume that you know enough about such matters for us not to dwell on them in this book. However, generally speaking, these assumptions cannot be made in the case of the special methods that attach to neuroscientific work.

Thus each of the following chapters reveals much of the content of neuroscientific work, focusing on topics sometimes quite different from the ones we have studied so far, and also its distinctive methods. In each, however, the emphasis is on the contribution that this area of neuroscience makes to our understanding of cognition, and we advise you to keep this firmly in mind as you encounter material that may appear as new and unfamiliar.

In Chapter 13, Gianna Cocchini introduces cognitive neuropsychology, a key discipline at the boundary of cognitive psychology and neuroscience. Cognitive neuropsychology is a field with a long history, having developed from a mix of research-related and clinically-related concerns. The chapter introduces the topic in the context of its historical development, indicating the evolution and multiplicity of the goals of neuropsychology, and the varied techniques that neuropsychologists use to understand a patient's impairment and its possible significance for models of cognition.

To sense some of the complexity involved, you might like to imagine trying to work out the information-processing characteristics of your PC or Mac from an analysis of its errors. More often than not, it is extremely difficult to gain an understanding of information processing from observing just one (impaired) computer. Perhaps through observing many different computers, and systematically observing the patterns of impairment, one might ultimately begin to uncover some of the structure of their information processes. In spite of the inherent difficulties the strategy faces, this is the position of the cognitive neuropsychologist with regard to human information processing or cognition.

The difficulty of drawing inferences from patterns of impairment about the nature of cognitive processes is much reduced by making certain assumptions, and the author discusses a number of these – such as subtractivity, for example, and the central assumption of modularity (which is further discussed in Chapter 20). The chapter also details many of the techniques used by neuropsychologists, such as standardized tests and the use of experiments. The chapter deliberately avoids in-depth discussion of some techniques that are being increasingly used by neuropsychologists, however. Cognitive modelling and neuroimaging are frequently used in conjunction with neuropsychological findings, providing researchers with (hopefully) converging evidence from different sources. However, these techniques are detailed in Chapters 14 and 19, and so no involved discussion of them is offered here.

In Chapter 14, Ingrid Johnsrude and Olaf Hauk introduce another method that can rightly lay claim to being a significant growth area within neuroscience: neuroimaging. Neuroimaging is a rapidly developing field, with many techniques having been developed and many more in development. The authors describe a number of different techniques for imaging the brain's activity. Each depends on being able firstly to detect the activity of neurons by some means, and secondly to generate a measure of the activity. The measure can then be used to provide images displaying the spatial or temporal locus of neuronal activity and, perhaps more importantly, to support statistical analysis. The authors outline a number of different techniques in all, though they concentrate primarily on EEG, MEG, PET, and fMRI, outlining the limitations and advantages of each.

Much controversy still surrounds the use of neuroimaging evidence. As becomes apparent in the course of Chapter 14, the different means for measuring neuronal activity are limited. For example, EEG and MEG have good temporal but poor spatial resolution – they are better at determining when neuronal activity takes place than exactly where. The reverse pattern holds for PET and fMRI. Such limitations – and many more are discussed in the chapter – remind us that neuroimaging techniques are informative only to the extent that we are prepared to draw *inferences* from the data they provide to the nature of cognitive processes, and that the validity of these inferences depends on many assumptions being satisfied. One thing that all the techniques have in common is that they measure *physical* qualities of neuronal activity. But cognitive processes are usually cast in terms of *information processing*. How can we be sure that physical activity is the vestigial trace of cognitive activity; that the physical signs we observe correspond to cognition?

Of course, researchers have developed strategies to enhance the reliability of their inferences. By comparing the images deriving from the performance of tasks that differ in just one component, for example, researchers can be more confident that the physical activity they see corresponds to the cognitive activity required by that task component. However, you can perhaps sense the difficulties involved if you imagine trying to produce physical images of a computer's activity. One could perhaps produce images of your PC or Mac's electrical or magnetic activity during a word-processing task, and you might find that particular regions were active in certain tasks and not others. But it is hard to see how the images alone would tell you exactly what information processing the computer was performing. To draw those kinds of inference we would need some prior knowledge of the information-processing characteristics of the task – at the very least we would need a good information-processing model – and it would help if we knew quite a bit already about the computer's architecture. In psychology, neuroimaging data are similarly likely to prove useful to the extent that we already possess good cognitive and brain models. This is not to say that neuroimaging data are not useful – they can help us discriminate between models that make different predictions concerning the locus and type of cognitive processes involved in certain tasks, and they may continue to throw up new data implying the involvement of processes in tasks that were previously unmodelled. Nevertheless, just *how* important neuroimaging data will prove for understanding cognition remains to be seen.

Chapter 15 by David Okai, Angus Gellatly, and Anthony David introduces the relatively new discipline of cognitive neuropsychiatry. Like its sister discipline of cognitive neuropsychology, it attempts to integrate findings from neuroscience with those from cognitive psychology in seeking a deeper understanding of disordered cognition and behaviour. The key difference is that the disorders that are the subject matter of this discipline – dementia, Parkinson's disease, and obsessive-compulsive disorder, to name just some of those considered in Chapter 15 – are ones traditionally studied (and treated) by medical science. However, as the authors make abundantly clear, successful explanations of these conditions are likely to contain an irreducible psychological or cognitive component.

The chapter begins by introducing cognitive neuropsychiatry. It describes the case of Mr E, a 56-year-old diagnosed with a Korsakoff-like amnesia (see Chapter 8). What is clear from this case study is the importance of developing a rounded understanding of individuals such as Mr E in order to properly diagnose and treat their conditions. Thus, it is vital to understand someone's pre-illness history, their social and psychological functioning, and their own understanding and perception of their symptoms and illness. Herein lies a general moral for the relation of cognitive neuropsychiatry to cognitive psychology. A full understanding of disorders and their relation to normal cognition also requires an understanding of a wide range of factors – social, cognitive, behavioural, and biological. As the chapter progresses we see that researchers are already postulating the roles played by these different factors.

The chapter next considers brain injury, and the likely dysfunction associated with damage to various lobes of the brain. Here we see for the first time issues that will recur later in the chapter. For example, we learn that damage to the occipital lobes, the location of primary visual cortex and of visual maps of the seen world, can cause visual hallucinations. Later in the chapter we learn also that visual hallucinations occur in Alzheimer's dementia, and dementia with Lewy bodies.

The remainder of the chapter is devoted to detailed consideration of different disorders, focusing first on dementias of varying kinds, as well as the related syndrome of Mild Cognitive Impairment. The chapter then turns to psychosis, schizophrenia, and delusions, and then to depression, disinhibitory disorders such as obsessive-compulsive disorder and hoarding, impulse control disorders, and finally conversion disorders, a mysterious set of disorders with no neurological cause (once termed hysteria). Throughout the chapter, the authors present an overview of what is known about these disorders, intertwining results from detailed neurological study with behavioural and cognitive studies. For example, we learn that in depression, individuals may show enhanced processing of and attention to 'sad' stimuli (e.g. words such as 'sad') and that this is related to an enhanced neural response in the dorsolateral prefrontal cortex. We learn that the dopamine system is linked to hoarding behaviour in Parkinson's disease, and it has been proposed that dopamine may cause patients to overestimate the intrinsic value of their possessions, resulting in hoarding.

The interweaving of biological and cognitive factors in explaining and understanding psychiatric disorders raises an important issue, which the chapter cannot fully resolve. Indeed, we already see some hint of this in the preceding discussion of hoarding behaviour. Where there are both biological and cognitive components to a disorder, it is tempting to think of the biological ones as more basic, as more fundamental, and as more likely to be causal. However, there remains the possibility that cognitive explanations have a degree of autonomy – that they cannot be reduced to biological explanations, and that it is not right to think of the biological level of explanation as causal, and the cognitive level as descriptive (see also the similar discussion in Chapter 20). It is possible, for example, that dopamine may cause overestimation of the value of possessions, but in theory it is also possible that overestimation might cause elevated levels of dopamine. To further complicate matters, it is possible that both are caused by some as yet unidentified factor, and that despite being correlated they are not causally related. These are of course theoretical possibilities, and the value of carefully designed experimental investigations is that they can begin to tease such possibilities apart. At this stage in the development of cognitive neuropsychiatry, however, as with the other disciplines of neuroscience that make contributions to cognitive psychology, we should be careful in how we interpret claims that imply a clearly understood relation between the biological and cognitive realms.

This page intentionally left blank

CHAPTER 13

COGNITIVE NEUROPSYCHOLOGY

Gianna Cocchini

Understanding the organization of mental activity and its relationship with the brain is one of the most fascinating challenges for science. Our brain constantly processes stimuli seeking to integrate this information alongside our pre-existing knowledge of the world, our emotions, and our goals. We may not be aware of this happening, despite the fact that, to some extent, this information affects our behaviour and our cognitive functioning.

Cognitive psychology has provided an incredible contribution toward its aim of a better understanding of mental processes. Scientists have formulated theories, refined methods and assumptions, and experimentally investigated their hypotheses with healthy volunteers. **Cognitive neuropsychology** has a very similar aim, but it approaches the general question from a different perspective. Rather than investigating mental processes by looking directly at their normal functioning, cognitive neuropsychology focuses on cases where these processes have broken down, trying to use such evidence to draw conclusions about the nature of 'normal' functioning.

After an initial introduction to the concept of cognitive neuropsychology and a brief historical overview of this discipline, this chapter will consider a number of crucial assumptions and methodological issues in cognitive neuropsychology. Examples of different syndromes and related studies will be used to clarify and illustrate concepts and methods of investigation. These will also allow you to explore different contexts, increasing your general knowledge of different syndromes and their theoretical interpretations. A final section will address more pragmatic aspects of neuropsychological investigations in clinical populations.

1 INTRODUCTION: THE LESION METHOD AND COGNITIVE NEUROPSYCHOLOGY

The lesion method represents one way of studying cognitive processes. Its basic assumption is that if a brain area plays a crucial role in a cognitive task, and this area is damaged, then performance of this cognitive task should be impaired. By carefully observing what happens when a system is damaged, we can understand more about the mechanisms of this system. Let us consider a classic example often used in textbooks and lectures. If someone wants to understand how a car works, then one possible way is to remove, or damage, some of its components and then observe what happens. Some functions will still work as before, whereas others may not. For example, if you damage the gear box you may observe that it is still possible to start the engine, all lights and indicators work fine, and if the car is placed on a slope it can still be moved and stopped. However, you soon realize that no gear can be engaged and that the car will not move on a flat surface despite the accelerator being pressed down. In other words, it is not only important to notice what is wrong but also what is still working, so that you can infer the normal function of the damaged component, its relationship with other parts of the car (for example, the accelerator), and its relative independence of other components (for example, functioning of the lights). Similarly, if we observe a change in cognitive performance after a brain lesion, we may be able to infer the previous normal function of the now damaged structure and its possible relationship with other mental processes.

The lesion method can be quite accurate with non-human animals as researchers can produce very circumscribed lesions (setting to one side the ethical questions raised by such work). However, with brain damage in humans, cognitive neuropsychologists have to rely on naturally occurring lesions, the extent and location of which are outside of the researcher's control. For this reason, some scientists have been quite critical about this method; their view is well summarized in this quotation by Marin *et al.* (1976): 'What can you possibly learn about the way a car works … by pounding it with a sledgehammer!' (p.868). In some cases this consideration is unfortunately quite right. Some patients suffer from extensive brain damage or multiple lesions and show very general cognitive impairment. Therefore, little can be added to our knowledge of mental processes by observing these cases. Moreover, the functioning of other unimpaired (or spared) neural regions that are connected to the damaged brain area might be altered after the injury, either because they are deprived of their neural input or because their normal synaptic connections are modified. For example, damage to some sub-cortical structures, such as the thalamus, or damage to fibres that connect brain areas, such as the corpus callosum, may reduce or affect input to other cortical areas, and the deficits may mimic impairments observed following damage of these specific areas. This may contribute to a confusing pattern of anatomical and cognitive impairment, leading to incorrect or simply uninteresting findings as suggested by Marin.

We will discuss possible criticisms of the lesion method and its assumptions later in this chapter. However, it is important to remember that researchers are not inexperienced scientists with no previous knowledge of the brain and cognitive processes. Interpretations of the cognitive symptoms (that is, the pattern of observable manifestations) are based on pre-existing knowledge of the brain and mental processes, and researchers' hypotheses and experiments are guided by theoretical concepts and cognitive models. As suggested by McCloskey (2001), complex systems, such as mental processes, 'can reveal their inner workings more clearly when they are malfunctioning [as in the example of our car] than when they are running smoothly' (p.594). Therefore, the expert scientific observation of behaviour following brain damage has provided an incredible contribution during the last two centuries (see the following section on historical milestones). While one should be aware of its possible limitations, the lesion method remains a valid and fruitful way to investigate mental processes.

Neuropsychology and cognitive neuropsychology are two similar and complementary scientific approaches, both founded on the lesion method, and both sharing an interest in pathological functioning of mental processes. However, the aims, methods, and final hypotheses of these two approaches are slightly different. The neuropsychologist attempts to draw conclusions about the function of different brain areas, explaining the patterns of impaired and intact cognitive performance in terms of damage to one or more of the brain areas and cerebral networks. Here, the researcher is focused on the relationship between cognitive processes and underlying anatomical substrates. The cognitive neuropsychologist is more interested in understanding and explaining the patterns of impaired and intact cognitive performance in terms of damage to one or more of the components of a theory of cognition, and in turn drawing conclusions about normal (healthy) cognitive processes. The final interest of the cognitive neuropsychologist is to fractionate a cognitive system and understand the role of each component of that system and its relationship with others (Ellis and Young, 1988; Coltheart, 1986). Determining which anatomical area underlies a specific process *per se* is of relatively little interest to the cognitive neuropsychologist (though of course of great interest to the neuropsychologist).

It is important not to overstate the contrast between the approaches of neuropsychology and cognitive neuropsychology, however. They share and contribute to similar literature, and sometimes it is not easy to distinguish between them. I prefer to consider them as different research approaches, as they are guided by different research questions, which implies some differences in the methods and techniques adopted.

The famous case of HM, the amnesic patient described by Milner (Scoville and Milner, 1957; Milner *et al.*, 1968), serves as a good example to clarify the slightly different approaches of neuropsychology and cognitive neuropsychology. In an attempt to treat a severe form of epilepsy, HM's hippocampi (located in the medial temporal lobes) were surgically removed (see Figure 13.2 for a diagram of some key anatomical and functional brain areas). The patient was left with profound amnesia. He could not remember new information (anterograde amnesia) and he had difficulty recalling events that had occurred in the last three years before the operation (retrograde amnesia). This case revealed two fundamental aspects of memory. The first was the relationship between the medial temporal lobes and memory; the second was that long-term memory is not a unitary cognitive system, and that memory traces undergo a consolidation process. Both aspects are clearly of great relevance to our understanding of human memory; however, while the approach of neuropsychology would be more focused on the anatomical relationship between hippocampi and memory, the approach of cognitive neuropsychology would be focused more on the different cognitive components of short- and long-term memory and the consolidation phases of memory traces. Box 13.1 discusses amnesia in more detail, focusing on the fascinating issue of amnesic patients' confabulations or false memories.

The study of limb apraxia offers another interesting example. Limb apraxia is defined as a difficulty in using objects or performing gestures, despite patients' spared ability to move their limbs. According to Liepmann's (1900) model, right and left anterior (frontal) areas of the brain are involved in the actual production of a gesture, whereas the mental representation of correct gestures relies on anatomical structures located in more posterior areas such as the left parietal lobe. Based on this model, Heilman *et al.* (1982) conducted a study to investigate the relationship between specific brain areas and type of apraxia. If a lesion occurs in more anterior areas of the brain, patients should show a difficulty in performing a gesture or using an object (for example, a hammer). However, their ability to recognize correct pantomime (the miming of tool use, such as miming brushing one's hair) or correct use of an object should remain preserved, as their ability to mentally represent a gesture is spared. Conversely, damage in left posterior areas should affect representation of gestures. The patients should not be able to correctly recognize or associate gestures and objects and, due to lack of input to the more anterior areas, these patients should also fail to correctly reproduce a gesture. Based on anatomical lesion (anterior or posterior), Heilman and colleagues considered two groups of patients who were asked to: (i) pantomime gestures and (ii) identify correct gestures performed by the examiner. Patients with posterior lesions failed both types of task, while patients with anterior damage mainly failed the task involving production of the gesture. Their findings were therefore in line with the anatomical role of anterior and posterior areas of the brain suggested by Liepmann.

Now let us consider the model shown in Figure 13.3 (Cubelli *et al.*, 2000), an example of a recent cognitive model accounting for use of objects and gestures. The different components of this model do not refer to different brain areas; thus, understanding of the impaired and normal cognitive functions can only be based on

BOX 13.1 Confabulation and false memory

Some brain-damaged patients show difficulties in recalling information previously encoded: they are amnesics. When questioned about past events, some of these patients may look puzzled, frustrated, or confused while they try to retrieve that specific information. However, other amnesic patients may provide florid recollections of false information, although it is not their intention to lie. For example, if asked about a recent holiday, they may provide detailed descriptions of events that never happened in places where they have never been. In some cases, their confabulations may grow to such an extent that they may start to behave in line with their invented stories. In accordance with their beliefs, they may report specific events and plans that seem initially to be realistic.

The mechanisms underlying the phenomenon of spontaneous confabulations are still debatable. Schnider and colleagues (Schnider *et al.*, 1996; Schnider and Ptak, 1999; Schnider *et al.*, 2000) have suggested that a possible explanation may rely on spontaneous confabulators' tendency to confuse the time at which information was last encountered. Indeed, it is interesting to note that the content of these invented events can be quite easily traced back to some aspects of actual experience. According to these authors, patients showing spontaneous confabulations may be unable to successfully repress old and irrelevant memories, which then interfere with the recollection. For example, after brain damage a patient who used to frequently travel by plane may believe that he or she is a pilot. In other words, Schnider and colleagues believe that spontaneous confabulators show a temporal confusion as they confuse ongoing reality with the past.

To test this hypothesis, Schnider and Ptak (1999) devised an interesting paradigm where healthy controls and amnesic patients (both confabulators and non-confabulating patients) were asked to perform a continuous recognition task. The participants were presented with a list of meaningful pictures and they had to report whether a picture was previously presented in the *current* trial (recurrent picture). Patients were asked to perform the same task for three additional trials (one immediately after the first, one after 5 minutes, and one after 30 minutes). The recurrent pictures in one trial were also presented in the following trials, but only once (i.e. as non-recurrent pictures). The authors' hypothesis was that, since confabulators show a difficulty in suppressing previously acquired information, they would show an increment of false positive responses from trial to trial. Therefore, compared to non-confabulating amnesics, confabulators would falsely recognize targets that had appeared once in a particular trial as recurrent stimuli, because these targets were displayed a few times previously in preceding trials. In other words, spontaneous confabulators would confuse stimuli presented in the current trial with those displayed in previous trials.

The results showed that both amnesic groups of patients performed worse than healthy controls at detecting recurrent stimuli (Figure 13.1a) due to their memory deficits; however, only spontaneous confabulators showed an evident increment of false alarms over the four trials (Figure 13.1b), supporting the authors' hypothesis of a time confusion of acquired information. Schnider and Ptak's (1999) interpretation is particularly interesting as it implies that some aspects of amnesia, such as confabulations, may be linked to activation of too many memory traces and not necessarily to a lack of traces.

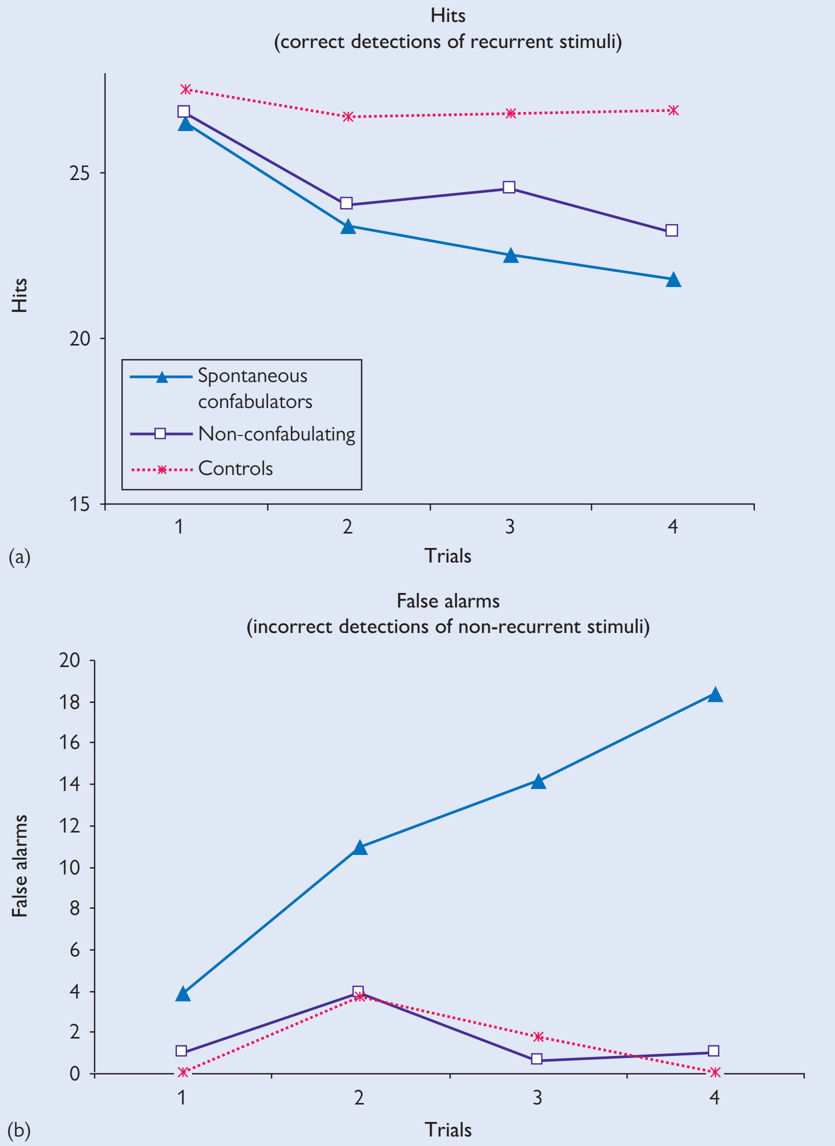


FIGURE 13.1 Hits and false alarms for each trial.

damaged/spared components of the model, not on damaged/spared brain areas. The authors themselves emphasize that this model is 'entirely based on cognitive concepts' in order to avoid 'confusion between cognitive and anatomical categories within the model' (p.146). In this context, research is not guided by the site of brain lesions but by the potential patterns of cognitive impairment that could further refine the model.

The important issue for our purposes is to define aims and methods in order to interpret findings within the cognitive neuropsychology approach. Cognitive neuropsychology is interested in patterns of intact and acquired pathological mental processes, and this chapter will investigate its assumptions, methods, and techniques.

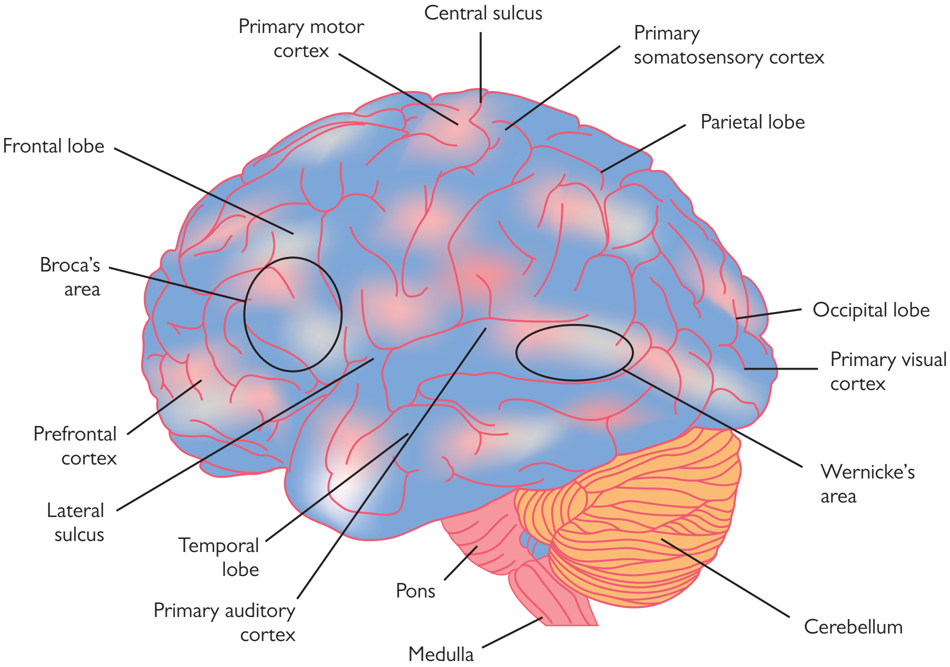

Primary motor cortex
Central sulcus
Primary somatosensory cortex
Parietal lobe
Frontal lobe
Broca's area
Occipital lobe
Primary visual cortex
Prefrontal cortex
Wernicke's area
Lateral sulcus
Temporal lobe
Pons
Primary auditory cortex
Medulla
Cerebellum

FIGURE 13.2 Main anatomical structures and functional areas of human brain.

SUMMARY OF SECTION 1

- Observation of malfunctioning mental processes may facilitate understanding of normal underlying cognitive mechanisms.
- The lesion method assumes that if a brain area plays a crucial role in a cognitive task, and this area is damaged, then performance of this cognitive task should be impaired.
- Cognitive neuropsychology is based on the lesion method and its aim is the understanding of normal (healthy) cognitive processes by observing the patterns of impaired and intact cognitive components after brain damage.

2 HISTORICAL MILESTONES

Despite the development of cognitive neuropsychology having occurred only in the last 40 years or so, it is important to consider the historical background that led to the birth of this discipline. During the second half of the 19th century some neurologists started to describe selective cognitive impairments in some patients who suffered from varying types of brain damage. Harlow (1848) described a patient, the famous Phineas Gage, who showed a dramatic change of personality following a traumatic head injury that damaged a specific portion of the prefrontal lobes. Before the accident Gage was described by friends and relatives as a highly capable and reliable man; however, after the brain lesion Gage became irritable, irreverent, socially inappropriate, and unable to accomplish any plan. A few years later Paul Broca

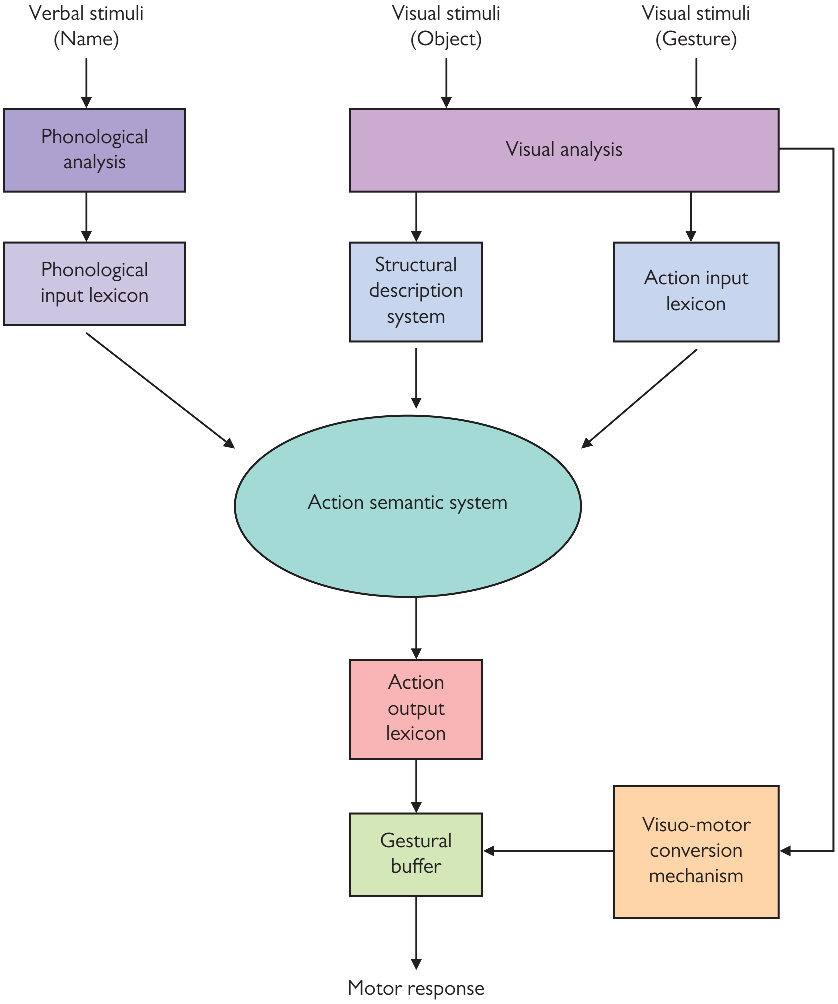

Verbal stimuli (Name)
Visual stimuli (Object)
Visual stimuli (Gesture)
Phonological analysis
Visual analysis
Phonological input lexicon
Structural description system
Action input lexicon
Action semantic system
Action output lexicon
Gestural buffer
Visuo-motor conversion mechanism
Motor response

FIGURE 13.3 Cognitive model of limb apraxia. Redrawn from Cubelli *et al.*, 2000.

(1861), in line with previous work by Bouilland (Selnes, 2001), described a patient showing an impairment of language mainly limited to productive aspects of language. Despite relatively preserved verbal comprehension, his speaking difficulties progressively deteriorated to such an extent that he could only say the word 'tan' before his death. A post-mortem examination revealed damage in the left frontal lobe (see Figure 13.2). In the same period, Wernicke (1874) reported a small group of patients showing a lesion in the left temporal lobe (see Figure 13.2) whose fluent language production contrasted with their poor verbal comprehension (Wernicke's aphasia). These cases represented the first attempt to study and interpret cognitive processes and their anatomical relationships. These scientists started to represent cognitive processes with diagrams containing boxes and arrows representing different cognitive processes and the links between them. The scientists, also known as diagram-makers or mapmakers, were the first to attempt to fractionate cognitive processes (i.e. to divide up the processes into putative component parts). By means of their diagrams, they aimed to illustrate sub-components of a specific process and their functional connections, and by extension, to predict possible malfunctions (deficits) of a process. A crucial aspect of the diagram-makers' approach was the anatomical correspondence of the sub-components, which later became the weakest and most criticized aspect. A classical example is Lichtheim's (1885) diagram of language, initially based on Wernicke's anatomo-clinical model of language function. One of the earlier versions of Lichtheim's model is represented in Figure 13.4. In this example, capital letters represented sub-components (also identified as 'centres') and lowercase letters represented input and output, while numbers represented points of malfunction resulting from different types of language impairment. For example, considering point 1, a malfunction of M (the motor word-representation centre) would result in a difficulty in language production (Broca's aphasia). It is important to note that Lichtheim's centres were not identified with single anatomical areas but rather with more distributed neural networks.

However, as mentioned earlier, the anatomical localization of the cognitive sub-components represented one of the weakest aspects of these diagrams.

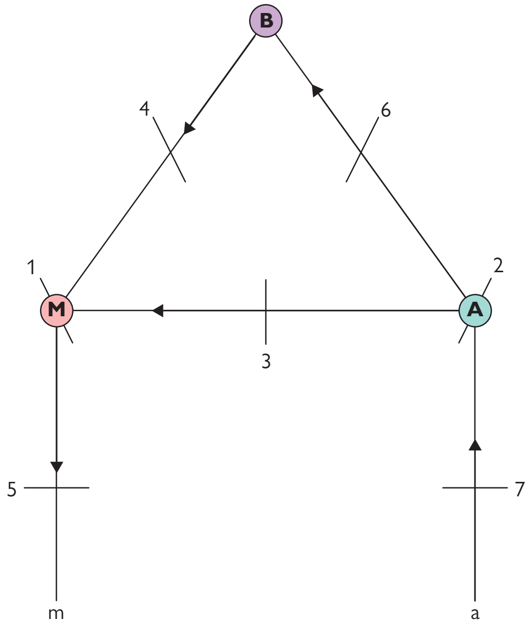


FIGURE 13.4 Lichtheim's diagram of language. A = auditory word-representation centre; B = concept elaboration centre; M = motor-word representation centre; a = auditory input; m = output (speech); numbers 1–7 represent points of malfunction. Redrawn from Lichtheim, 1885.

Indeed, revisions of anatomical evidence of crucial cases were strongly criticized by several neurologists (e.g. Marie, 1906) and, in the absence of supporting evidence, were considered by most of the scientific community as mere speculations. Following this period, studies of groups of patients with batteries of psychometric tests were preferred to single-case observations. The methodological rigours of group studies appeared to be the best way to collect solid empirical data, and were considered a robust method of investigation. However, in the 1960s there was a resurgence of interest in single-case studies, probably because, according to Vallar (1999), researchers realized that cognitive architecture is too complex and that group studies should not represent the sole method of investigation.

Some case descriptions and analyses of single-case studies had a profound influence that would result in a new approach to studying mental processes. Geshwind (1965) revisited a series of clinical studies to define his concept of a disconnection syndrome; Milner (1966) reported the famous case of HM who, as described above in Section 1, showed severe memory disorders following a lobotomy of the temporal lobes to treat a severe form of epilepsy. Luria (1966; 1973) reconsidered the concept of localization of functions proposed by the diagram-makers, suggesting that some functions are complex and required a multifaceted interaction of different brain areas. Marshall and Newcombe (1966) described patient GR, who showed a pattern of reading impairment mainly characterized by semantic errors. This case and the debate about new language models represented a theoretical and empirical context for the birth of the modern approach of cognitive neuropsychology. In the preface to the classical book *Deep Dyslexia*, considered a milestone of this new approach, Coltheart and collaborators (1980) emphasized that 'We shall only progress beyond the classical "diagram-makers" if we take as seriously as they did the requirement that our conjectures about disordered systems should be expressed sufficiently formally that counter-examples can be recognised as such' (Coltheart *et al.*, 1980; p.viii). Within this new theoretical approach, brain localization of the deficit was of little (if any) interest, whereas the cognitive impairment became 'an experimental condition (induced by nature) suitable for probing specific hypotheses relevant to normal cognitive processing mechanisms' (Selnes, 2001; p.38).

3 ASSUMPTIONS IN COGNITIVE NEUROPSYCHOLOGY

3.1 Modularity

One of the core assumptions of cognitive neuropsychology is modularity: a specific assumption about how our mind is organized. A detailed description and theoretical discussion of modularity can be found in Chapter 20. In the present chapter, I summarize the fundamental aspects for cognitive neuropsychology.

The development of the modularity concept mainly derived from philosophers and linguists, such as Fodor (1983) and Chomsky (1980), but was also informed by the work of psychologists such as David Marr (1976). Their work has provided an invaluable contribution to our understanding of how most of our mental processes are organized around cognitive processors called modules. For each higher cognitive process we should be able to identify a set of simpler but independent processes. For example, according to the model presented in Figure 13.3, a cognitive function, such as recognizing a gesture (e.g. clapping hands) and its meaning, will consist of different stages: encoding of visual perception; visual analysis of the gesture; recognition of the gesture; and finally access of semantic information to retrieve the meaning of the gesture. Marr (1976) also added that independence of processes is fundamental, because were the processes not designed in this way, a small change in one process would have consequences in many others, making it impossible for us to understand the structure of the human mind.

However, Fodor himself (2000) warned about the intrinsic risks of considering all cognitive processes as modular. He suggested that only input systems (i.e. systems devoted to encoding and recognizing incoming information) are modular, while central systems (i.e. higher cognitive functions such as fixation of beliefs, expectations) are not modular. Modularity of the input system is well represented in the 'box and arrow' cognitive models (as in the one represented in Figure 13.3), where each module (or box) transforms information from one stage to the other. But what is a module, and what are its characteristic features?

In *The Modularity of Mind*, Fodor (1983) detailed nine features of modules, which are reported in more detail in Chapter 20. First, modules are usually *domain specific*, meaning that each module can only serve one particular type of input. For example, the module processing the colour of a stimulus is not involved in processing the shape of the same stimulus. Second, modules are usually *mandatory*, meaning that they are unstoppable and independent from voluntary control. In other words, we cannot avoid processing a noise that falls within hearing range, or recognizing a visual stimulus. This is an important aspect for Fodor, as he considered this feature as the criterion by which we distinguish between input systems (which are modular) and central systems (which are not modular). The well-known Stroop effect (see Box 2.2 in Chapter 2) occurs because irrelevant information (e.g. meaning of the colour word) is automatically processed, despite the task requiring the participant to report the colour of the letters forming the word (for example, the word 'RED' written in green ink). The irrelevant information (the word 'RED' in our example) interferes with the relevant information (the colour green in our example). Good performance depends on participants' ability to inhibit the irrelevant information that their brain has processed. The Stroop effect is more evident in some patients with brain damage in prefrontal areas, as they have great difficulties in inhibiting the irrelevant information and this seriously interferes with their performance (Perret, 1974).

The third feature of modules described by Fodor is that they are usually *innate and genetically determined*. Fourth, modules are usually *informationally encapsulated*, which means that a module carries out its own process regardless of what happens in the rest of the cognitive system. This is one of the most crucial aspects of modularity. Visual illusions are classical examples of the concept of informational encapsulation. Indeed, visual illusions remain vivid even when we are well aware of the illusion. The Ponzo illusion (discussed in Chapter 20) and the well-known Müller–Lyer illusion (see Figure 13.5) are good examples. Despite knowing that the horizontal lines of the Müller–Lyer illusion are equally long, the one with the arrowheads pointing outwards is perceived as longer than the other. Similarly, if you gently push your eyeball, you will experience an apparent movement of your surroundings even if you know that you are voluntarily moving your eyeball.

The fifth feature is that modules are usually *fast*. This is linked to the features of being mandatory and automatic. Sixth, modules are usually *neurologically wired*, which means that modules underlie similar neural architecture. The other three properties that Fodor proposed as characteristic of modules are outlined in Chapter 20.

As clarified by Coltheart (1999), Fodor did not consider any of the above features as necessary conditions to define a cognitive system as modular. Therefore we can still consider a cognitive system as modular even if it does not possess *all* nine of these features. For example, some modules are involved in acquired cognitive processes, such as reading or writing, and they are not innate. According to Coltheart, only domain specificity should be considered as a crucial feature for modularity, while for every other feature 'it is an empirical question whether or not the proposed module has that feature' (p.115).

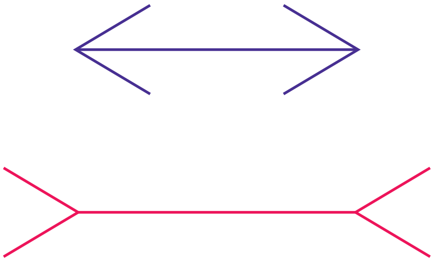


FIGURE 13.5 Müller-Lyer visual illusion. Despite the two horizontal lines being identical lengths, the one above appears considerably shorter than the one below.

In the context of cognitive neuropsychology, the modularity assumption has represented a powerful way to describe and conceptualize cognitive functions. Indeed, since the modularity assumption suggests that a cognitive function (e.g. object recognition) can be broken down into simpler components, then damage to one of these components will selectively impair a part of this cognitive process and will result in an observable cognitive deficit (e.g. apperceptive agnosia). Note that in this context I am referring to cognitive, not anatomical, damage. However, as underlined by Humphreys and colleagues (2006), some manifestations may result from complex underlying mechanisms and not from damage to a single, well-defined module. They may result from interactions between different modules that can be modified by adaptation or by strategies adopted after the brain lesion has occurred.

3.2 Subtractivity and transparency

Two other important assumptions in cognitive neuropsychology should be considered. These are subtractivity and transparency. According to the assumption of subtractivity (or fractionation), an impaired system is just the same as a normal one except that certain modules or processors of that normal system have been damaged. Therefore it is possible to conceptually 'fractionate' normal cognitive systems. For example, to be able to name an object, we need to perform a visual object recognition, then retrieve the semantic information associated with this object, then retrieve the phonological information by accessing the phonological output lexicon and buffer, and finally say the name aloud. A cognitive neuropsychologist will observe different types of errors depending on which of the above modules is affected by brain damage.

The concept of subtractivity is deeply embedded in another important assumption of cognitive neuropsychology, which is transparency. This assumption implies that the pattern of symptoms observed following a brain lesion reflects underlying damage of specific components pre-existing the brain injury. This is a fundamental assumption of the lesion method, as cognitive neuropsychologists aim to study pathological conditions in order to make inferences about the structure of normal cognitive processes. In this context, it is important to establish the actual causal link between the observed symptoms and the brain lesion. In other words, a patient may show weak performance in some tasks even pre-morbidly (i.e. before brain damage). This possibility of individual variability is something we will return to in Section 4.2 where we consider possible limitations of single and group studies.

Although it is widely accepted that an adult brain is not capable of considerable functional reorganization after a brain injury, such damage may lead, to some extent, to the development of more or less conscious strategies (which we will discuss in more detail in Section 5.1) that induce brain-damaged patients to process information in a different way. In some cases, patients' performance may mimic a neural reorganization, which may not necessarily reflect the generation of new neural networks but the strengthening (or reduced inhibition) of previously latent links. These cases are still consistent with the assumptions of transparency and subtractivity as they are the expression of latent, not new, processes. Blindsight and phantom limb phenomena may offer good examples. Blindsight is a phenomenon observed in some patients with a blind visual field due to damage of the primary visual area or of the geniculostriate pathway (the cortical pathway carrying visual information from the lateral geniculate nuclei of the thalamus into the primary visual area; see Figure 13.6). Surprisingly, given they report blindness in their visual field, blindsight patients seem to be able to perform some simple visual tasks, though they are unaware of having processed such visual information. Interpretations of this phenomenon have been quite controversial; however, some authors (Weiskrantz *et al.*, 1974; Weiskrantz, 1986) suggested that this phenomenon may depend on the spared sub-cortical visual system. Indeed, a small amount (about 10 per cent) of the visual fibres do not follow the geniculostriate pathway, instead connecting with sub-cortical structures, such as the superior colliculi (see Figure 13.6), which could mediate some aspects of localization of visual stimuli in blindsight (Rafal *et al.*, 1990). Under normal conditions, this information is of negligible consequence for our conscious experience of visual stimuli, being relevant for the regulation of eye movements. However, following particular brain damage of the main cortical pathway, in some patients this pathway becomes an important source of information and it guides some unconscious responses to visual stimuli. Similarly, the phantom limb phenomenon, which occurs after amputation of a portion of a limb, has been recently interpreted as the result of pre-morbid latent brain connections. In these cases the brain is not damaged, but somatosensory cortical areas are no longer receiving afferent information from the corresponding part of the amputated limb. Interestingly, if a tactile stimulus is applied on circumscribed areas of the patient's face, the touch is experienced on their lost hand (e.g. Ramachandran *et al.*, 1995). Despite initial suggestions that this was due to new cortical networks developed after the amputation (e.g. Basso and Pizzamiglio, 1999), it has now been suggested that the phantom limb phenomenon may depend on the existence of latent pathways between adjacent areas in the primary somatosensory cortex (as in the case of hands and face) (Ramachandran *et al.*, 1995). These latent connections play a negligible role in intact brains, but after brain damage their activity may be either more evident or strengthened, as possible inhibitory input from the damaged area is now reduced or halted (Gazzaniga *et al.*, 2002). The point here is that some functional reorganization after brain damage may be possible even in adult brains; however, this may simply be due to a change in strengthening of pre-existing connections, whereas

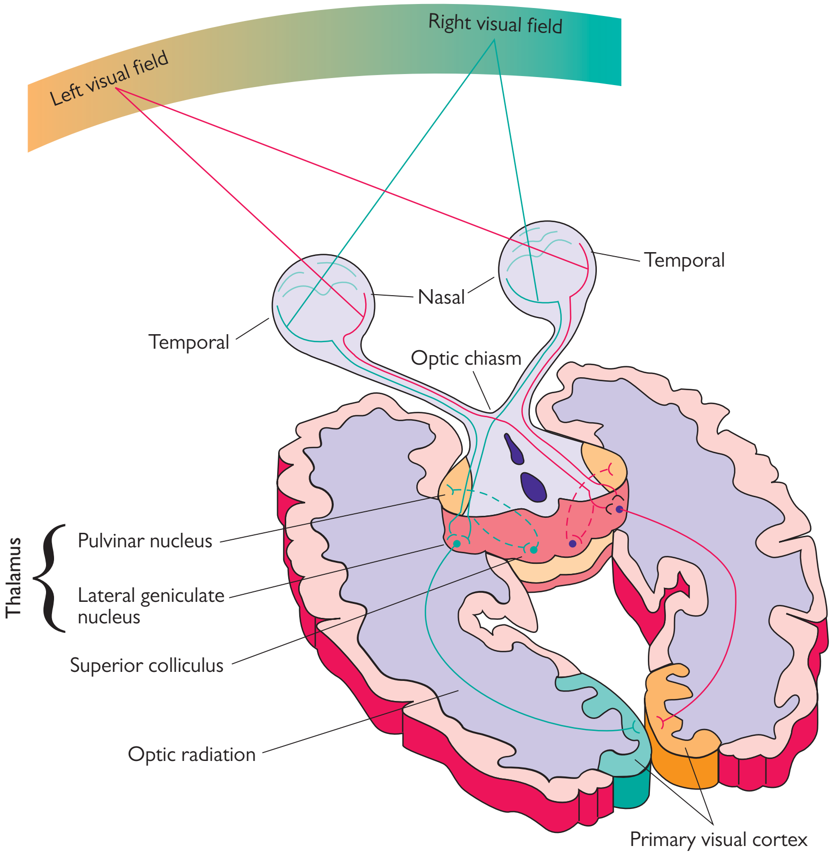


FIGURE 13.6 Visual pathways from retina to primary cortex and sub-cortical areas.

regeneration of neuronal connections is still quite unlikely (Della Sala and Bates, 1996).

For our purposes, these phenomena are interesting as they illustrate the concept that, despite the limited functional reorganization of mental processes that can be observed in some circumstances, generally it can be reasonably assumed that the pattern of symptoms observed following a brain lesion reflects damage of specific cognitive components that pre-existed that brain lesion.

3.3 Anatomical and functional syndromes

A syndrome is a term used to define a particular recurrent association of a cluster of deficits (also called symptoms). For example, **neglect** syndrome refers to a patient's difficulty with detection of stimuli from all modalities presented in the contralesional (i.e. opposite side to the brain lesion) side of the space or of the body. In Figure 13.7, for example, it can be seen that the patient suffering from neglect, having experienced right brain damage, experiences difficulty detecting or reproducing objects in their left visual field.

The term 'syndrome' is mainly useful for clinicians and researchers to enable quick communication of information regarding cognitive, sensory, and motor impairments, and the likely site of the lesion in a patient or group of patients. For example, if we refer again to the neglect syndrome, we may expect the cognitive impairment described above to be manifest in association with extinction (which is the inability to detect a contralesional stimulus when presented simultaneously with another one in the same side as the brain lesion), and anosognosia (lack of awareness

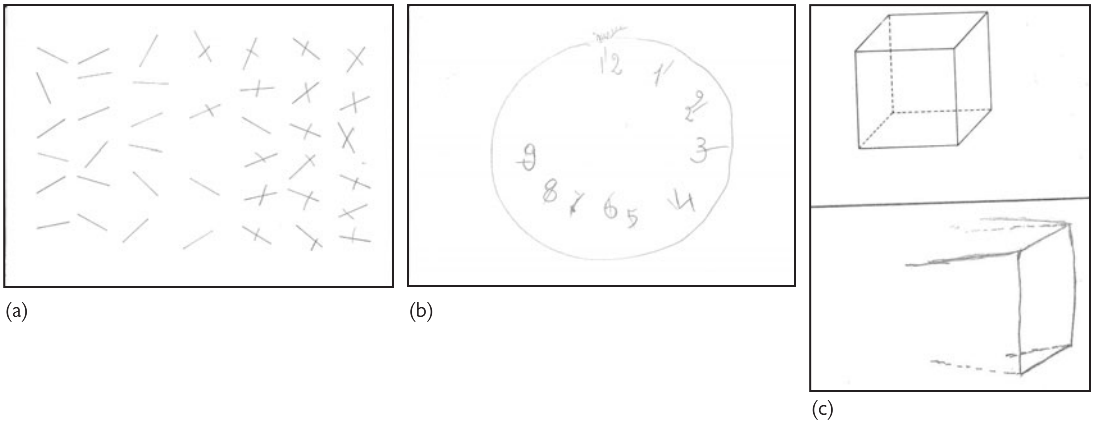


FIGURE 13.7 Examples of a right-brain-damaged patient's performance on some neglect tests: a) line cancellation test; b) drawing a clock from memory; c) copy of a drawing.

of illness). We may also predict, with some approximation, that the patients are likely to show a brain lesion in the right inferior parietal lobe and that they are likely to show left weakness or hemiplegia (inability to move one half of the body). Similarly, the classification of language disorders, as in Broca's (difficulty in verbal expression) or Wernicke's (mainly difficulties in verbal comprehension) aphasia, has been based on associations of specific symptoms.

However, even if the concept of a syndrome appears relatively simple and straightforward, it has important implications and its use within the cognitive neuropsychology framework has been a subject of debate over the past 30 years. An obvious problem is that symptoms in a specific syndrome may be quite heterogeneous. This could be due to the fact that a cognitive process may consist of several sub-systems that can be independently 'lesioned'. If so, we may observe great variability in performance across different patients, where each patient may show a slightly different pattern of symptoms. Here, adopting too stringent a criterion to define a specific syndrome could result in a proliferation of many highly specific syndromes, a possibility that has led some authors to stress the risk of ending up with as many syndromes as patients (e.g. Patterson *et al.*, 1985; Caramazza, 1986; Ellis, 1987). Partially for this reason, Morton and Patterson (1980) and, with an even more extreme point of view, Caramazza (1986) rejected the idea that there are specific syndromes that can be identified by a set of specific tests. According to them, patients are unique, and the use of syndrome terminology is only useful to facilitate communication, being of little interest for theoretical purposes. However, if patients are unique, then replication of findings with other patients and by other research teams would be very difficult, if not impossible.

Another group of researchers have considered the concept of syndrome from a different point of view. In 1966, Marshall and Newcombe (see also Marshall and Newcombe, 1973) described a patient, GR, who has been the subject of much debate about the definition of deep dyslexia and, more generally, about the concept of syndrome in cognitive neuropsychology. Following left hemisphere brain damage, GR presented with long-lasting language disorders, mainly consisting of (though not limited to) reading impairment. In particular, this patient committed a very high percentage of semantic errors (for example, he read 'dinner' as 'food'), and he showed an effect of syntactic class. GR could not read non-words (such as 'wep') and his reading performance was at its worst with function words (such as 'for'; often responding 'I don't know'), less compromised with verbs (such as 'play'), less again with adjectives (such as 'tall'), and finally better (though still below 50 per cent) with nouns. GR's performance suggested a main impairment of the semantic route of the dual-route model (further described in Section 4.1). However, other patients diagnosed with deep dyslexia have shown additional difficulties, such as reading regular words (e.g. 'if'), suggesting a further malfunction of the phonological route of this model (which would explain why regular words could not be read by applying the letter-to-sound rules). GR also presented with several other deficits, such as impairment of spontaneous speech,

BOX 13.2 Implicit processing in neglect syndrome

Patients suffering from unilateral neglect tend to ignore stimuli located on the contralesional side (i.e. opposite side to the lesion, usually the left). For example, patients may ignore objects, or even people, located on the contralesional side; they may bump into doorways, neglect to dress one side of their body, or comb their hair on one side only; finally, they may even omit the details of one side of a mental image when describing it. However, some information is not completely lost and there is evidence that some implicit processing of these stimuli occurs.

In 1988, Marshall and Halligan described a patient who became a classic study in the neglect literature. The authors described a right-brain-damaged patient (PS) showing a severe form of left neglect. PS was presented with two line drawings of a house. The houses were vertically aligned and identical, except that the left side of one house was on fire (burning house). The patient was asked to describe the pictures, and PS claimed that they were identical houses across several presentations. However, when asked to indicate in which house she would prefer to live, she consistently chose the non-burning house while claiming that the task was silly 'because they're the same' (p.766). When the pair of houses were presented with flames on the right (non-neglected) side, PS immediately spotted the difference. This patient showed a clear dissociation between explicit (when she stated that the stimuli were the same) and implicit (when she chose the non-burning house) processing of the stimuli.

Further evidence of implicit processing of stimuli explicitly ignored was demonstrated by Vallar and colleagues (1994), whose patients were shown a series of pairs of animal drawings aligned vertically. Within each pair, one drawing was a chimeric figure that consisted of the front of one animal attached to the back of another animal. The right side of each pair showed the same animal. Patients were asked to decide whether the two pictures were the same or different. They mistakenly reported the pairs as similar in the vast majority of the cases. However, when patients were given a verbal cue to the animal on the left side and asked to point to the figure that most resembled it, patients tended to either consciously identify the neglected half or to point to the chimeric figure.

These findings suggest that verbal cueing facilitates processing of neglected stimuli and, for some patients, is even sufficient for the recognition of previously neglected data. But why? Ellis and collaborators (1996; Beschin *et al.*, 1997) pointed out an interesting interpretation involving a specific role of working memory. These authors hypothesized that visuo-spatial working memory may represent a strong approach for interpreting some aspects of neglect syndrome. According to them, perceptual input (including that which was 'neglected') reaches long-term memory, activating semantic traces and forming an interpretation of that information to be passed onto working memory, where conscious acknowledgement of stimuli occurs. In neglect patients, the portion of the working memory for the contralesional hemifield is not (or is less) activated. However, the consciously neglected information may 'reach a sufficiently high level of activation to exceed the threshold necessary for the patient to give non-random responses . . . and allow for implicit processing' (Ellis *et al.*, 1996; p.77). Within this theoretical framework, additional information, as with the verbal cue in Vallar and colleagues' study, may sufficiently enhance the level of activation of information that it can pass into the visuo-spatial working memory.

agrammatism, agraphia, and errors in auditory word-picture matching. Looking at all of these deficits, it can be confusing to define the characteristic features of a syndrome. What then defines a specific syndrome? Coltheart (Coltheart, 1980, 2001; see also Marshall and Newcombe, 1973) suggested that, starting from the most striking symptom (in the case of deep dyslexia, the most striking symptom would be evidence of semantic errors), if a cluster of other symptoms are recurrently associated, then we should consider this pattern of symptoms as a syndrome, or 'symptom-complex' (p.23). Within this context, the absence of some deficits is fundamental; for example, it is fundamental that deep-dyslexics are still able to name objects (i.e. they are not severely anomic), whereas other symptoms are irrelevant for the definition of deep dyslexia.

Other syndromes can be defined in a similar way. Patients showing unilateral inattention (also called neglect; see our earlier discussion) can present with a quite heterogeneous pattern of deficits. Indeed, inattention can be found in one modality (e.g. for visual stimuli) but not in others (e.g. tactile stimuli); moreover some, but not all, patients also show visual extinction, unawareness, and general depletion of attentional resources. However, the most striking symptom is inattention for stimuli placed in the contralesional hemifield (Figure 13.7) and this, with the pattern of deficits recurrently associated, defines this syndrome. Box 13.2 discusses the fascinating phenomenon of neglect in more detail.

Association of symptoms may therefore have important functional implications; if so, a specific pattern of symptoms can be considered as a prerequisite to define a specific syndrome. However, in some cases a cluster of symptoms may be recurrently associated but their co-presence is not functionally and hierarchically ordered. So, while deep dyslexia can provide an example of a functional syndrome, in other cases co-occurrence of symptoms may be due to anatomical proximity. In these cases, brain damage, especially if relatively large, is likely to impair different cognitive processes, resulting in a particular association of deficits that do not represent a common underlying functional mechanism. Deficits may be associated simply because the anatomical substrates of different cognitive processes are physically close or adjacent and the brain damage may encompass all of them. A typical example of an anatomical syndrome is **Gerstmann's syndrome**, diagnosed when the following cluster of symptoms are observed: finger agnosia (difficulty in identifying a finger when touched); acalculia (difficulty in performing arithmetic calculations); right-left disorientation; and pure agraphia (difficulty in writing). Despite some attempts to identify a functional link between these symptoms, such as distortion of the body schema (Gerstmann, 1930), it is now generally accepted that the association of these symptoms is due to the proximity of the anatomical substrates of different cognitive systems that can be damaged by a single brain lesion in the left parietal lobe (Poeck and Orgass, 1966).

SUMMARY OF SECTION 3

- Core assumptions of cognitive neuropsychology consist of specific assumptions about how our mind is organized.
- The main assumptions are that: (i) our mental processes are organized around cognitive processors called modules, which have specific characteristic features (modularity); (ii) a damaged system is just the same as a normal one but with certain modules or processing systems damaged; therefore it is possible to fractionate mental processes (subtractivity); and (iii) the pattern of symptoms observed following a brain lesion reflects underlying damage of specific cognitive components that existed prior to the brain damage (transparency).

- The concept of syndrome is quite controversial. Some authors reject this concept in favour of individual description of single patient symptoms. Others consider that it is possible to reasonably identify a syndrome as a specific pattern of recurrent association of symptoms.
- Associations of symptoms can be due to damage of the same cognitive function (functional syndrome). In some cases, association of symptoms can be due to anatomical proximity (anatomical syndrome).

4 METHODS IN COGNITIVE NEUROPSYCHOLOGY

4.1 Association, dissociation, and double dissociation

We discussed in the previous section that some symptoms may be commonly associated following specific brain damage and that, despite some exceptions, this fact usually implies some type of functional relationship. Therefore, we might be tempted to conclude that, if a patient shows symptom A and symptom B, then these two symptoms are the result of damage to a common cognitive process. However, such an association can also result from simultaneous damage to different cognitive processes, and any inference of a common cognitive mechanism may be incorrect or misleading. Coltheart (2001) offers an example to illustrate this point. The dual-route theory (see Chapter 6) can be used to interpret reading and writing impairments. This theory proposes two possible ways that one can read aloud: the first is by applying letter-to-sound rules to the printed string (non-lexical route), whereas by the alternative route (the lexical route) the word accesses the orthographic input lexicon where it is recognized and then linked to the correct pronunciation in the phonological lexicon. Should the orthographic input lexicon (which is part of the lexical route) be impaired, then patients would still be able to read non-words (for example 'choe') and regular words (for example 'cup'), as they can do so by applying the letter-to-sound rules. However, the patients would fail to read irregular words (for example 'shoe') as correct reading of these words does not rely on the letter-to-sound rules, and attempts to read irregular words by the non-lexical route would result in regularization errors (i.e. reading 'shoe' rhyming with 'toe'). It is quite common to observe patients showing impaired ability to read and write irregular words, but spared abilities with non-words and regular words in reading and writing (called surface **dyslexia** and surface **dysgraphia**, respectively). Based on this association, we could infer that there is a unitary orthographic lexicon for input (reading) and output (writing) processes. The hypothesis of two separated orthographic lexicons, one for input and one for output, would be seen as a 'non-economical' redundancy. However, if a second patient, or group of patients, shows poor performance on one task (e.g. writing irregular words) but performs normally on another task (e.g. reading irregular words), then the previous hypothesis of a unitary mechanism runs into difficulty. Indeed, a selective failure in writing irregular words would imply a deficit of the output orthographic lexicon, whereas a selective impairment of reading irregular words would imply a deficit of the input orthographic lexicon. In this scenario, association of deficits may not necessarily imply a unitary cognitive module and, as in the example above, surface dyslexia and dysgraphia may be due to a concurrent damage of different cognitive modules.

Selective impairment in one task, but preserved performance in another task, is known as dissociation.

The use of dissociations, rather than associations, represents a more powerful method for cognitive neuropsychology, as the presence of a dissociation suggests the presence of different underlying mechanisms and, in turn, allows us to fractionate a cognitive process. A classical example of dissociation is that between short- and long-term memory. Some patients with memory impairment show poor performance in tests assessing long-term memory, but show no problems in tasks requiring them to remember a few items for a short period of time (short-term memory). In 1970, Baddeley and Warrington presented a study with patients suffering from memory impairment and a control group of healthy participants. Participants were asked to perform a serial recall task with a list of ten words. Typically in such tasks, the words presented in the last three or four positions in the list are recalled sooner and with more ease (the **recency effect**) than words presented earlier in the list (the **primacy effect**). Most psychologists agree that the recency effect depends on those last words still being present in the short-term memory storage, or working memory, whereas the earlier items have already been transferred to the long-term memory storage (Atkinson and Shiffrin, 1971; see also Chapter 9 for a detailed description of the multi-store theory of memory). Baddeley and Warrington's control group (see Figure 13.8) showed the expected performance, where more recent words were recalled better than earlier words. Amnesic patients showed a similar pattern of results for later words (recency effect). They did also show a primary effect, since at the beginning of the task they seemed to have sufficient resources to transfer some information to long-term memory; nevertheless, their performance for earlier words was significantly worse than that of the controls. These findings, supported by the vast literature on memory impairment, represented a clear dissociation between short- and long-term memory, and had profound theoretical implications.

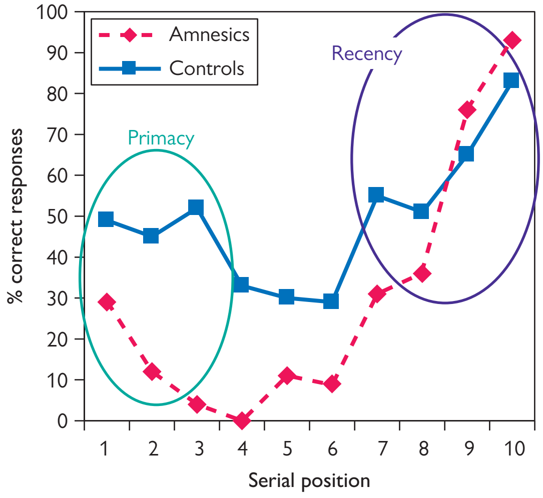


FIGURE 13.8 Controls' and amnesics' serial position effect in recalling a list of ten words (Baddeley and Warrington, 1970).

Though dissociation is a crucial method, it is important to be able to compare opposite patterns of dissociations (double dissociations) in order to avoid potentially incorrect interpretations. If a patient shows a deficit in task A but not in task B, we cannot directly infer that these two tasks test different underlying cognitive mechanisms. Task A and B may assess the same process, but task A may be placing a higher cognitive demand than task B on the same underlying mechanism (that is, the comparison may introduce a resource artefact). However, if a second patient shows the opposite dissociation, that is, good performance in task A and poor performance in task B, then we have a double dissociation (two patients showing opposite patterns of data). Here, the previous criticism that there may be a possible bias due to a resource artefact is no longer valid and we can be more confident in assuming that tasks A and B are assessing two different cognitive mechanisms (see also Chapter 20 for a discussion of double dissociations). Despite some criticisms (e.g. Charter and Ganis, 1991; Robertson *et al.*, 1993), the double dissociation of single cases is a prevailing method for cognitive neuropsychologists, and findings by these means have greatly enhanced our understanding of cognitive processes. For example, going back to the memory system, the distinction between short- and long-term memory is accepted, but a strong assumption of the modal model (one of the earlier interpretations of memory functions; Atkinson and Shiffrin, 1971) was that information first entered short-term storage and then was transferred into long-term memory storage. In other words, this theory accounts for a selective impairment of long-term memory, as in Baddeley and Warrington's (1970) study, but not for the opposite dissociation, that is, a selective impairment of short-term memory. When patients showing memory deficits limited to short-term memory started to be described (e.g. Warrington and Shallice, 1969) the modal model ran into serious difficulties. The convergence of all of these studies provided evidence for a double dissociation between short-term and long-term memory, suggesting a direct (maybe additional) route for long-term memory (Ellis *et al.*, 1996). This contributed to the development of new theories that considered different links and relationships between different aspects of memory (see Chapter 9).

While the use of double dissociations is a valuable and widely accepted method, it is important to discuss some issues about dissociations raised by Shallice in 1988 in his now classic book, *From Neuropsychology to Mental Structure*. Let us consider the case where task A is performed in the normal range, while task B is performed very poorly and, if compared with norms, performance is clearly pathological. Shallice called this a 'classical dissociation'. In a double classical dissociation, two patients show opposite performances in two different tasks. For example, a patient may be unable to pantomime the correct gesture associated with an object on view, but has no problem in performing the correct gesture if asked verbally (i.e. the name of the object is given), whereas a second patient may show the opposite dissociation. However, following brain damage, classical dissociations are quite unusual, and another type of dissociation, called 'strong dissociation', is much more common with clinical samples. In this case, both patients show a pathological performance on *both* tasks, but in opposite patterns of performance when the two are compared. For example, patients showing surface dyslexia are likely to show greater difficulties in reading irregular words (for example, only 40 per cent of the irregular words are correctly read) than regular words (for example, 85 per cent of the regular words are correctly read). However, healthy control performance for both types of words is at ceiling (at or near 100 per cent); therefore in the case of these patients, their reading of regular words, despite being much better than their reading of irregular words, is also at a pathological level. Phonological patients, on the other hand, seem to show the opposite pattern. For example, Job and Sartori (1984) described a phonological dyslexic patient, Leonardo, who read 45 per cent of irregular verbs correctly, but only 18 per cent of regular verbs. Again, in this case reading of both types of verbs is pathological, but reading of regular verbs is clearly worse than reading of irregular verbs. Evidence of a strong (rather than a classical) dissociation does not diminish the importance of this dissociation, which suggests existence of different mechanisms.

Similarly, as we have seen, it is quite common that neglect patients omit many, if not all, visual stimuli on the contralesional side (neglect side). But they may also fail to detect a few stimuli on the ipsilesional side (the non-neglected side), whereas healthy control performance on both sides is at ceiling. This can be due to the fact that similar cognitive impairments may have similar symptoms. In this last example of neglect, we can assume that a patient may suffer from a general attentional disorder, maybe reduced vigilance, in addition to the more specific lateralized deficit, neglect (Robertson *et al.*, 1994). Here, the more general attentional disorder results in omissions with no particular spatial organization (i.e. also in the ipsilesional side). We cannot exclude the possibility that some of the contralesional omissions could also be due, to some extent, to the general non-spatial attentional disorder, but the considerable difference between ipsilesional and contralesional sides strongly supports the hypothesis of a lateralized attentional disorder.

However, Shallice (1988) has pointed out a possible risk with strong dissociations. Tests offer an occasion to 'measure' how well a cognitive mechanism is working; nevertheless, not all tasks are equally affected by a brain lesion, and the relationship between performance on specific tasks and the resources necessary to accomplish these tasks may be different. For example, an easy task may require relatively little resource to reach ceiling performance, and any further increment in resources may not result in a further performance improvement. On the other hand, performance on a more difficult task may benefit from additional increments in resources (see Figure 13.9 for performance/

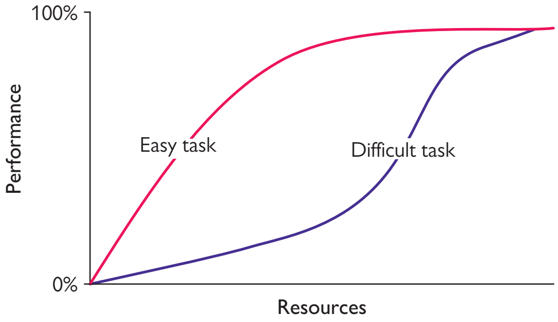


FIGURE 13.9 Two types of performance/resource curves. Redrawn from Shallice, 1988.

resource curves for such hypothetical easy and difficult tasks). Now let us consider how brain damage may affect resource availability and hence task performance. Suppose that patient 1 performs better on task A than task B, and performance of both tasks is better than that of patient 2, who performs better on task B than task A. This result may depend on different performance/resource curves linked to the two tasks. This is illustrated in Figure 13.10, where we can see that increments in performance in task A initially require considerable increments in resources, whereas performance in task B increases smoothly with increases in the available resources. Following brain damage we may observe an overall depletion of resources and this may penalize task B considerably more, resulting in the strong double dissociation just discussed. In this case, however, the strong double dissociation would be insufficient evidence of the existence of two mechanisms, as the resource artefact cannot be ruled out.

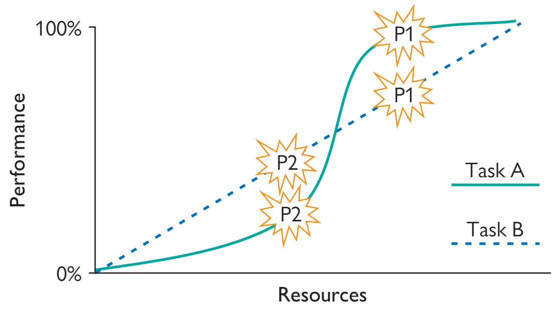


FIGURE 13.10 Possible performance/resource curves of two tasks (A and B) to illustrate resource artefact in strong dissociation. P1 and P2 indicate performance of patient 1 and patient 2 in both tasks. Redrawn from Shallice, 1988.

Other factors may bias patients' performance, such as a lack of comprehension of the task (task-demand artefact; Shallice, 1988) or different strategies used by the patients. For example, poor verbal comprehension may affect understanding of task instructions on tests not assessing language impairment. These may appear as quite intuitive problems but they are tremendously important when carrying out research with brain-damaged patients who are likely to present with a number of cognitive deficits that may interfere with their performance over and above the primary cognitive function impaired. These factors may represent a source of potential confounding variables, and it is therefore important to establish the effect of any deficit on different tasks.

Despite these considerations and the potential pitfalls of dissociations (see Ward, 2006), classical and strong dissociations remain an important and effective method to investigate cognitive processes. In cases where the above criticisms and artefacts are overcome, if we observe a double dissociation then we can reasonably infer that two different mechanisms are underlying the tasks. In turn, this implies that different cognitive modules are damaged and this evidence will contribute to the researchers' ability to refine or devise theories representing 'a reasonable interpretation of data' (Coltheart, 2001; p.15).

4.2 Group studies versus single-case studies

The decision to carry out single-case or group studies has profound implications for the issues discussed earlier in this chapter.

Group studies consist of the investigation of a group of patients all meeting a certain criterion (for example, having a brain lesion in a similar area or showing a particular syndrome, etc.). These types of studies represent a valuable method for investigating cognitive processes and their relationship with anatomical substrates; they are also fundamental in responding to questions about the incidence and frequency of specific deficits. Moreover, group studies are generally less exposed to criticism such as poor replicability, which is an important aspect of modern scientific research (Popper, 1959), or difficulties in using the studies to support generalizations. Robertson *et al.* (1993) also considered group studies to be a better way to investigate cognitive modularity. Patients' performance is usually compared with that of a group of control participants (usually healthy volunteers or other patients not showing the crucial symptom or specific brain lesion). The control group is usually comparable (similar) to the patient group across a series of variables (for example, age, gender, years of formal education, etc.). An example of a group study is reported by Baddeley and Warrington (1970) who compared a clinical population of amnesics with a group of healthy volunteers (see Section 4.1 in this chapter). Alternatively, a second clinical group with similar characteristics to the experimental sample but not showing a specific deficit can represent the control group (e.g. Schnider and Ptak, 1999).

A crucial assumption of group studies is that the clinical sample is homogeneous with respect to a specific feature, and patients are representative of a clinical picture. The assumption of homogeneity is not only in relation to spared cognitive processes, which can be considered as a generally valid assumption (but see Section 5 of this chapter); the homogeneity is also assumed for the impaired function, which may be more problematic to accept *a priori*. Indeed, some authors have pointed out that, especially with brain-damaged patients, group performance may be a sort of artefact that does not reflect the performance of each individual within the same sample (e.g. McCloskey, 2001; Caramazza and Badecker, 1989; Sokol *et al.*, 1991). This criticism has been well explained by Banich (2004). Suppose we decided to test nine brain-damaged patients and nine healthy participants using three different tasks (see Table 13.1). Looking at Table 13.1, we can see from the overall averages obtained in the three tests that patients performed worse than controls on all tests, but we also notice that performance across all three tests was very similar within each group. Based on this last observation we may be inclined to conclude that the group of patients did not show a difference across the three tests. However, if we look at patients' individual data we can identify three subgroups. Patients 1–3 performed particularly poorly in task 1, while their performance is similar to those of controls in tasks 2 and 3. Similarly, patients 4–6 showed a poor performance only on task 2, and patients 7–9 showed a poor performance only on task 3. Therefore, the average score of the patient group does not represent the performance of any individual or subgroup, making it difficult to draw any conclusion about the possible cognitive process that is impaired in these patients.

TABLE 13.1 Example of potential distortion of overall performance pattern when group data are averaged (Banich, 2004)

Patient	**Performance (% accuracy)**		
	Task 1	Task 2	Task 3
1	20	75	70
2	25	70	80
3	30	80	75
4	75	25	75
5	80	30	70
6	70	20	80
7	75	80	25
8	70	75	30
9	80	70	20
Average	58.3	58.3	58.3
Controls' average	78	80	83

The functional architecture of mental processes (i.e. the organization of cognitive modules and their relationships) is highly complex and in some cases the single-case approach may be a more efficacious method to investigate specific cognitive models and related theories. The main criticism of group studies consists of the fact that it is very difficult to select a homogeneous group that is still representative of the population under examination. As pointed out by Vallar (1999): 'A cerebral lesion, the localisation and size of which reflect anatomo-physiological factors, may disrupt more than one component of the system. Accordingly, if a group of patients is selected on the basis of the side or site of the lesion, it remains possible that patients with heterogeneous behavioural deficits, and,

✱ ACTIVITY 13.1

Limb apraxia has been defined as an impairment of gesture processing not due to motor impairment *per se*. Usually this syndrome is observed following lesions of the left hemisphere. Looking at the cognitive model in Figure 13.3, we can identify two possible routes used to process information and perform an action. One route (the lexical route) involves accessing the semantic system, whereas the other route (the non-lexical route) does not access the semantic system. In this second route visual information is processed by means of visuo-motor conversion mechanisms that allow imitation. Now, suppose you want to find evidence for this two-route distinction using the lesion method (that is, testing brain-damaged patients). What type of experiment would you devise? Try to identify:

1. Hypotheses and predictions – What type of behaviour/deficit would you expect to observe if either of these routes is impaired? In other words, which type of dissociation would you expect to observe?
2. Inclusion criteria – Would you select a specific type of patient based on site of the lesion?
3. Exclusion criteria – Could any associated deficit (e.g. aphasia, agnosia, etc. . . .) represent a confounding variable for your experiment?
4. Crucial tasks and conditions – Which is the crucial aspect of the lexical and of the non-lexical route? Identification of this crucial variable will guide the main structure of your experiment.
5. Dependent variable/s and establish how you will analyse data.

A few guidelines and points for further discussion:

- Consider meaningful (e.g. pantomime the use of a hammer) and meaningless gestures and the relevance of semantic access.
- Consider gesture production versus gesture imitation and when access to the semantic system is crucial.
- Since the non-lexical route converts visual information into motor representations, you may need to consider visual (rather than verbal) input also for the task assessing the lexical route.
- Consider which type of error you may expect for malfunction of either ‘box’ or ‘link between boxes’ (i.e. the arrows).
- Make sure that associated deficits, such as visual agnosia, will not bias the findings. How will you control for these possible confounding variables?

Also, note that the two routes converge in the gestural buffer (which holds information on line and translates abstract formats into sequences of motor commands). A deficit at this level may affect final performance of both tasks that depend on processing in the lexical route, and tasks that depend on processing in the non-lexical route.

by implication, damage to discrete functions, are pooled together. Also a selection on the basis of the presence or absence of a clinical syndrome (e.g., Broca’s aphasia, or spatial hemineglect) is open to this criticism, as patients with non-homogeneous cognitive disorders might be included.’ (p.121).

Caramazza (1986), who has strongly supported the use of single-case studies, pointed out that in the attempt to look at global findings, interesting individual results might pass unnoticed. Rare double dissociations between two patients, or even within the same patient, as in Beschin *et al.*’s (2000) patient who showed left perceptual neglect *and* right representational neglect for mental images, may be hidden in the averaged group performance. In this context it is not very important whether many patients will show a similar pattern of symptoms; what is really relevant here is that these findings can further fractionate a cognitive process and refine a theoretical interpretation. As pointed out by Ward (2006), the single-case method is not a prerogative of cognitive neuropsychology, as it is also used in other scientific environments, and the rarity of some single cases does not reduce their scientific relevance. For example, Ward noted that the discovery of the primate Lucy by Johanson in 1974 has represented a crucial case for modern theories about the evolution of primates. Before this discovery, scientists believed that, from a phylogenetic point of view, brain enlargement had preceded the ability to walk upright. However, Lucy's skeleton was that of a woman able to walk upright despite her relatively small brain, and 'Johanson did not have to provide a group of Lucy for his findings to be acceptable to the scientific community' (Ward, 2006; p.85).

In conclusion, the choice between using single-case or group studies has been the subject of great debate and different research teams have reached quite extreme opinions. However, both methods have provided valuable contributions to the understanding of cognitive processes. Researchers need to be aware of the criticisms that each method is exposed to, and the related implications. The research goals should then guide selection of the most appropriate method.

Now that you have considered some of the key methodological features of work in cognitive neuropsychology, try Activity 13.1 and see how you would design a neuropsychological study.

SUMMARY OF SECTION 4

- Association of symptoms can guide definition of syndromes, which may be relevant to facilitate communication amongst researchers and clinicians and to define inclusion criteria in research studies. However, association of symptoms *per se* is not a powerful method to investigate cognitive functions.
- Dissociations and double dissociations consist of observing opposite patterns of performance on two different tasks and allow researchers and theorists to fractionate cognitive processes. Classical dissociation is observed when performance on one of two tasks is within a normal range, while strong dissociation is observed when performance on both tasks, despite being significantly different, is pathological.
- Group studies are carried out on relatively large samples and data are compared across participants. Groups are assumed to be homogeneous according to a specific characteristic crucial for experimental purposes. This assumption has some limitations, and appropriate selection of participants is one of the fundamental aspects of group studies.
- Single-case studies are conducted on a single patient showing a peculiar pattern of symptoms. These studies can offer the possibility of investigating rare and highly interesting cases from a theoretical point of view. The main limitation is the question of whether such findings generalize to healthy populations.

5 FURTHER CRUCIAL ISSUES FOR COGNITIVE NEUROPSYCHOLOGY

5.1 Compensatory strategies and related issues

Following brain damage, we may observe behaviour that cannot easily be explained as a direct result of the lesion. In these cases the pattern of symptoms may arise from some type of compensatory strategy. The literature on language disorders is rich with examples. In 1979, Butterworth described a patient (KC) who showed jargon aphasia, which consists of producing many neologisms in what otherwise appears to be fluent, but often incomprehensible, verbal communication. Butterworth noticed that KC's neologisms were either produced at the end of a sentence, or followed longer latencies compared to other words, and that the same or similar neologisms tended to be recurrent in a conversation. On these bases, it was suggested that KC's neologisms should be considered as 'filler' morphemes that the patient was generating to fill anomic gaps – locations in speech where KC was unable to select the correct noun. Therefore, Butterworth concluded that KC's speech features were mainly the result of a more or less conscious strategy adopted by the patient in response to his language difficulties.

Compensatory strategies are not just rare occurrences in single cases; indeed, some strategies can even become a recurrent adaptation for groups of patients, and may arise from pre-morbid processes (that is, processes already in place prior to brain damage). For example, Coltheart (2001) considered patients with pure alexia who have profound reading impairments, and may only be able to identify a whole word after having spelled it out letter-by-letter (so-called letter-by-letter reading). This reading procedure is not normally adopted by healthy adults, but we cannot assume that it has been introduced by the brain damage. This way of processing information exists for any healthy person, but it is underused as we normally adopt more efficient ways of reading. However, if a brain lesion impairs the more effective way of reading, then patients may adopt the 'underused process . . . using the normal functional architecture in an abnormal way' (Coltheart, 2001, p.10).

5.2 Pre-morbid pathological conditions and previous brain damage

As mentioned at the beginning of this chapter, cognitive neuropsychology assumes that the pattern of symptoms observed following a brain lesion reflects underlying damage to specific cognitive components that existed prior to the brain damage (assumption of transparency). Although this is often the case, in some instances this assumption may not be valid. A patient may have shown an atypical organization even before brain damage for various reasons. This could be due to developmental conditions (e.g. congenital blindness), psychiatric history, previous neurological diseases, or simply pre-morbid use of strategies, as seen in Levine and colleagues' (1982) patient who adopted a method of fast reading for years and attained a highly developed visual imagery ability, before suffering from a brain injury resulting in a peculiar dissociation between speech and reading.

This is a critical issue in many studies involving epileptic patients. These patients have usually suffered from epileptic seizures severe enough to induce patients and clinicians to consider neurosurgical intervention to avoid epileptic seizures spreading to other parts of the brain. As a result of this, study of these patients has given rise to the emblematic condition of disconnection syndrome, known as split-brain (Sperry *et al.*, 1969; Gazzaniga *et al.*, 2002). In these cases, a portion of the corpus callosum (the main commissural fibre that connects the same anatomical areas in the two hemispheres) is severed to avoid the spread of the epileptic seizure from one hemisphere to the other. Clearly these patients represent an incredibly interesting opportunity to investigate functions of each hemisphere (hemispheric asymmetry). However, as the patients may have suffered from very severe epileptic seizures since childhood we cannot rule out the possibility that their cognitive organization before neurosurgery was atypical. Similarly, epileptic patients have participated in another set of studies using the Wada test. This test, usually administered during the pre-surgical phase, consists of injecting a barbiturate into one of the two carotid arteries to temporarily inactivate one hemisphere. Again, this has represented an extremely interesting opportunity for neuropsychologists and cognitive neuropsychologists to investigate hemispheric specialization of particular cognitive processes, leading to crucial theoretical implications. However, as mentioned above, these patients' anatomical brain organization may be different from an intact brain, and cognitive organization may be atypical or dissimilar across individuals. If so, it would not be surprising to observe contrasting findings. For example, the Wada test has been used in a group of studies examining hemispheric specialization of awareness for motor impairment. During these studies patients showed motor impairment on the side of the body opposite to the anaesthetized hemisphere, and were questioned about their motor ability (e.g. 'Can you move your arm?') to evaluate whether they were aware of the deficit. Findings, summarized in Table 13.2, show quite a heterogeneous pattern of data (Cocchini *et al.*, 2009; Cocchini and Della Sala, 2010). Despite a general trend towards a crucial role of the right hemisphere in becoming aware (i.e. its inactivation would produce **anosognosia** – a condition in which patients are unaware of an impairment), the data across the studies are surprisingly heterogeneous. Some studies showed a very strong hemispheric effect (Gilmore *et al.*, 1992) while other studies found no hemispheric difference at all (Dywan *et al.*, 1995). Moreover, even within the same hemispheric condition, the frequency of anosognosia varied considerably. For example, when the left hemisphere was anaesthetized the percentage of patients showing anosognosia varied considerably across the studies from 0 per cent (Gilmore *et al.*, 1992) to 89 per cent (Durkin *et al.*, 1994). Despite the unlikelihood that a unique factor may explain these contrasting findings, we cannot disregard the possibility that such heterogeneous patterns of data may result from individual differences in cognitive organization, probably due to recurrent and long-lasting severe epileptic seizures in one part of the brain.

TABLE 13.2 Percentage of participants showing evidence of unawareness following suppression of activity of left and right hemisphere (Wada test). Modified from Cocchini and Della Sala, 2010

	Hemisphere anaesthetized		Number of participants
	RIGHT	LEFT	
Gilmore *et al.*, 1992	100	0	8
Buchtel *et al.*, 1992	92	59	48
Kaplan *et al.*, 1993	100	71	15
Durkin *et al.*, 1994	95	89	150*
	94	86	115*
Dywan *et al.*, 1995	66	66	83
Breier *et al.*, 1995	89	49	54
Carpenter *et al.*, 1995	87	38	31
Lu *et al.*, 1997	80	59	17

**The larger sample consisted of 150 participants (regardless of hemisphere dominance for language). The sub-group of 115 participants refers to dextrals with left-hemisphere dominance for language.*

Finally, it is also important to consider that brain damage can affect the pre-existing balance between different modules and that different brain lesions may have opposite effects, which can confound eventual findings. A dramatic example is reported by Vuilleumier and collaborators (1996). The authors described a patient who, following a first right parieto-occipital vascular lesion, showed a severe left neglect syndrome; the patient ignored stimuli presented on the left side. However, ten days after the lesion, the neglect vanished completely, immediately after a second vascular insult occurred in the left frontal lobe. The authors suggested that the first lesion caused an attentional imbalance, but that equilibrium was restored by the second lesion to a different brain site. Clearly brain damage does not improve cognitive processes; rather this case exemplifies how brain lesions may not only impair a particular process, but may also play an important role in rebalancing and modifying existing relationships between different modules.

5.3 Combination of deficits: the contribution and interaction of different deficits

A further complication, especially if patients show multiple lesions due to traumatic brain injuries, is that of the combined effect of different cognitive deficits, which may confuse or blur the clinical picture. For example, a few years ago with my collaborators (Cocchini *et al.*, 2002), we described a patient (NS) suffering from a severe form of anosognosia. Following bilateral brain damage, NS showed hemiplegia (i.e. paralysis of one side of the body) for his left limbs and a pattern of cognitive deficits including left personal neglect (i.e. lack of attention for the side of the body opposite to the brain lesion) and severe anterograde amnesia (difficulty in learning and memorizing information presented after the brain damage). Despite being confined all day to a wheelchair, NS denied his motor impairment (anosognosia for hemiplegia) to such an extent that he recurrently fell to the floor as he attempted to walk. After a series of experiments, we noticed that NS was able to acknowledge his hemiplegia when his attention was forced toward the neglected side of his body (asking him to attend to stimuli located on his left hand or foot), suggesting that personal neglect may have impeded 'discovery' of motor impairment. However, quite surprisingly, the effect was only temporary, and he would again deny his deficit a few minutes later. This was interesting as anosognosia had always been interpreted as a deficit in becoming aware, not *remaining* aware. In other words, if you painfully discover that your leg has been broken as a consequence of a recent fall, you don't need to attempt to walk each morning in order to remain aware that your leg is broken; you simply 'remember' it. If so, why did NS deny his impairment so soon after having acknowledged it? We interpreted this phenomenon as being due to his inability to memorize new information (anterograde amnesia), including the recent information related to his motor impairment. Therefore, we suggested that NS's chronic anosognosia was the result of the combined effect of personal neglect (which prevented NS from acknowledging his paralysis) and anterograde amnesia (which prevented him from remembering what was just acknowledged). This is quite a different way of considering chronic anosognosia to the traditional view: a chronic inability to become aware. Davies and collaborators (2005) pointed out the possibility that different mechanisms (and eventually the malfunction of different modules) may be responsible for some syndromes such as anosognosia. For this reason, it is important to provide what Shallice (1988) called 'the results of standard clinical baseline tests' (p.30), otherwise it may be difficult to compare results from different patients, and results interpreted within the framework of a theoretical context may actually reflect an impairment in another domain neglected in our standard assessment.

SUMMARY OF SECTION 5

- Compensatory strategies or adaptation may occur following brain damage, and this may affect the pattern of symptoms observed.
- The assumption of transparency may not always be valid, as pre-morbid mental organization may have been atypical in some cases.
- It is not unusual for the pattern of symptoms observed to be the result of a combination of different deficits, which may blur or complicate the clinical picture. It is important, therefore, to record patients' cognitive abilities on a series of tests in order to better identify and understand possible interactions or alternative explanations for findings.

6 MEASURING BEHAVIOURAL RESPONSES

The cognitive neuropsychologist is interested in the pattern of spared and impaired cognitive processes following brain damage. This pattern of data consists of detailed observation of patients' performance on a series of psychometric tests. These are behavioural data that then allow quantitative (test scores) but also qualitative (e.g. types of error) analyses.

'Pure' cases (i.e. those showing a selective cognitive impairment) are very rare and to rely only on these cases may not be the most efficient approach for cognitive neuropsychology. As mentioned earlier, it is particularly important to define a comprehensive profile of a patient's cognitive capacities, as many different reasons may account for impairment on the task in question. For example, a patient may not be able to perform the gesture necessary to use an object (e.g. a comb) placed on the table. This failure may be interpreted as limb apraxia. However, other reasons may account for a failure in performing such a task. For example, the patient may be either unable to actually see the object, which may have been displayed in a blind portion of the visual field (as in quadrantanopias or hemianopias), or they may have difficulties in directing attention towards the stimulus (as we have seen in attentional deficits and neglect). However, it is possible that the patient has no visual or attentional deficits, but is unable to recognize the object (agnosia). Finally, he or she may be unable to perform the task because of severe verbal comprehension deficits that prevent them understanding the task instructions.

A comprehensive battery of tests can provide the necessary information to understand whether patients show associated disorders that may affect their performance of a specific task. However, in order to explore all possible cognitive impairments, it may be necessary to use quite extensive and long-lasting batteries of tests. In these cases, the effects of tiredness or spontaneous recovery (if the assessment needs many test sessions) may bias the findings. This is a particularly relevant consideration for patients in the acute phase (i.e. soon after the brain damage) when a general spontaneous improvement is very likely to be observed within the first few days or weeks.

McCloskey (2001) also emphasized the importance of qualitative analyses, which address the types of error committed by the patients. Just think about the crucial role of semantic errors for the diagnosis of deep dyslexia. Going back to the example reported earlier of the patient failing to reproduce the correct gesture when a comb is placed on the table (main task), we may expect to observe different types of errors depending on the actual cognitive impairment. For example, if the patient fails the main task because he or she does not correctly recognize the object, then they may perform a pantomime corresponding to another object, and may be able to perform the correct pantomime if the name of the object is provided. However, if the poor performance is due to comprehension difficulties, the patient may not be able to understand the instruction of the task and we may expect a completely unrelated behaviour (e.g. the patient may nod or point to the comb, or merely seem puzzled).

Some information may already be available to the researcher from a preliminary assessment, which can include psychological, neurological, and radiological information. This type of information can guide further, more specific investigations. How we collect this specific information may vary considerably depending on whether we are conducting a group study or a single-case study (see Section 4.2 above).

With group studies, it is quite common to administer fixed batteries of tests to all patients. A very important issue in these types of studies is that of inclusion/exclusion criteria; in other words, how patients are selected for the study. For example, a study on apraxia may consider all patients showing: (i) brain damage in the left hemisphere; and (ii) evidence of difficulties in using objects (inclusion criteria). However, all patients showing severe language comprehension impairment on a preliminary test may be excluded from the study, as data may be unreliable (exclusion criterion). Selection of participants is an extremely important methodological step in cognitive neuropsychology, as an inaccurate selection may lead to incorrect conclusions.

For single-case studies the procedure is very different, as in these cases a patient is brought to the researcher's attention due to her/his peculiar characteristics. Usually a battery of tests and experiments will be tailored around the specific pattern of symptoms shown by this patient, and each test will be guided by specific clinical or theoretical questions.

7 CONCLUSIONS

The aim of cognitive neuropsychology is the understanding of normal (healthy) cognitive processes by observing the patterns of impaired and intact cognitive components after brain damage. The lesion method represents one way to fractionate cognitive processes, and to understand the role of specific subcomponents and their relationships. Concepts such as modularity, transparency, and subtractivity are the core assumptions of cognitive neuropsychology, as they represent a sort of guideline for the study of mental processes. Not only do these concepts guide our understanding, interpretation, and generalization of findings, but they also have profound implications for the choice of the most appropriate method to investigate a specific cognitive function.

We have considered the advantages, possible limitations, and criticisms of single-case or group studies. Although group studies are a valuable method of investigation, we have appreciated how single-case descriptions remain a crucial method for the cognitive neuropsychologist. Indeed, single-case studies are usually more appropriate for a detailed investigation of a specific pattern of symptoms, and it is of relatively little importance whether this occurrence is a rare condition or not, as long as these findings allow further insight into a cognitive process. For example, the description of a rare case showing a selective deficit of a cognitive component may lead to a further fractionation of a cognitive module. The rarity of the event does not diminish the relevance of the finding.

In this chapter we also discussed the relevance and limitations of associations and dissociations. Despite possible pitfalls, dissociations, and in particular double dissociations, offer the most powerful way to investigate and fractionate cognitive modules. In general, if we observe two patients with opposite performance on two different tasks, then we can reasonably suspect impairment of two different cognitive mechanisms.

That being said, the functional architecture of cognitive processes is complex and highly interactive. This means that cognitive processes are not isolated, and that the interpretation of findings can be further complicated by a number of crucial and critical issues. Indeed, after brain damage, a patient may compensate for the lack of an ability by using other strategies (albeit more or less consciously), which may blur our findings and lead us to misleading conclusions. Similarly, some patients may have pre-morbid conditions (such as previous brain damage or a congenital disease) that may have affected their cognitive organization. This may represent a problem for the important assumption of transparency and limit the generalization of related findings. All this does not necessarily mean that these patients should not be considered. On the contrary, they may still represent an interesting opportunity to investigate cognitive processes. However, the cognitive neuropsychologist will need to consider these factors and their possible implication in relation to specific findings.

Finally, the assessment is a delicate moment of the investigation. As I mentioned earlier, syndromes can be quite heterogeneous and identification of relevant symptoms may be complicated by various associated symptoms. It is important to obtain a general clinical and cognitive picture of each patient. Pure cases are not so frequent and often brain-damaged patients present a range of cognitive deficits; therefore, it is important to evaluate the implications and possible biases of different cognitive impairments on a cognitive task. Indeed, the functional architecture of mental processes is extremely complex and performance on a particular test may require several cognitive functions. Failure on a task may be determined by an impairment of any of the cognitive functions usually required. Therefore, data need to be considered in a wider scenario and possible confounding variables must be identified and addressed.

Although cognitive neuropsychology is a relatively recent discipline, our understanding of some cognitive functions, such as language, has grown considerably over the last few decades and detailed cognitive models have been developed to explain possible underlying mechanisms. Future research in cognitive neuropsychology will need to keep refining and challenging existing models. It should similarly address other cognitive processes, our understanding of which has moved forwards with more difficulty.

Acknowledgements

Special thanks to Max Coltheart for reading and providing valuable comments on a preliminary version of this chapter.

FURTHER READING

Shallice, T. (1988) *From Neuropsychology to Mental Structure*, Cambridge, Cambridge University Press. This classic book remains key reading for cognitive neuropsychologists. In it Shallice covers crucial methodological aspects and issues of data interpretation for research with brain-damaged patients.

Rapp, B. (2001) *The Handbook of Cognitive Neuropsychology*, Hove, Psychology Press. This book contains several chapters that critically consider theoretical and methodological issues for cognitive neuropsychology and specific syndromes.

Ward, J. (2006) *The Students' Guide to Cognitive Neuroscience*, Hove, Psychology Press. Despite addressing the wider field of neuroscience, Chapter 5 of this book covers a number of methodological issues overlapping with cognitive neuropsychology, while other chapters provide an overview of different cognitive impairments following brain damage.

For detailed and critical revisions of the historical phases of the study of cognitive processes following brain damage:

Selnes, O.A. (2001) 'A historical overview of contributions from the study of deficits', in Rapp, B. (ed.) *The Handbook of Cognitive Neuropsychology*, Hove, Psychology Press.

Shallice, T. (1988) 'From the diagram-makers to cognitive neuropsychology', in Shallice, T. *From Neuropsychology to Mental Structure*, Cambridge, Cambridge University Press.

For detailed information on psychometric tests and batteries:

Lezak, D.M., Howieson, D.B., Loring D.W., and Hannay, H.J. (2004) *Neuropsychological Assessment*, 4th edition, Oxford, Oxford University Press.

Strauss, E., Sherman, E.M.S., and Spreen, O. (2006) *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*, 3rd edition, Oxford, Oxford University Press.

REFERENCES

Atkinson, R.C. and Shiffrin, R.M. (1971) 'The control of short-term memory', *Scientific American*, vol.224, pp.82-90.

Baddeley, A.D. and Warrington, E.K. (1970). 'Amnesia and the distinction between long- and short-term memory', *Journal of Verbal Learning and Verbal Behavior*, vol.9, pp.176–189.

Banich, M.T. (2004) *Cognitive Neuroscience and Neuropsychology*, Boston, Houghton Mifflin Company.

Basso, A. and Pizzamiglio, G. (1999) 'Recovery of cerebral functions', in Denes, G., and Pizzamiglio, L. (eds.) *Handbook of Clinical and Experimental Neuropsychology*, Hove, Psychology Press.

Beschin, N., Basso, A., and Della Sala, S. (2000) 'Perceiving left and imaging right: dissociation in neglect', *Cortex*, vol.36, pp.401–14.

Beschin, N., Cocchini, G., Della Sala, S., and Logie, R. (1997) 'What the eyes perceive, the brain ignores: a case of pure unilateral representational neglect', *Cortex*, vol.33, pp.3–26.

Breier, J.L, Adair, J.C., Gold, M., Fennell, E.B., Gilmore, R.L., and Heilman, K.M. (1995) 'Dissociation of anosognosia for hemiplegia and aphasia during left-hemisphere anesthesia', *Neurology*, vol.45, pp.65–7.

Broca, P. (1861) 'Remarques sur le Siège de la Faculté du Langage Articulé, Suivies d'une Observation d'aphémie (perte de la parole)', *Bulletin de la Societeé Anatomique*, vol.6, pp.330–57.

Buchtel, H., Henry, T., and Abou-Khalil, B. (1992) 'Memory for neurological deficits during the intracarotid amytal procedure: a hemispheric difference', *Journal of Clinical Experimental Neuropsychology*, vol.14, pp.96–7.

Butterworth, B. (1979) 'Hesitation and the production of verbal paraphasias and neologisms in jargon aphasia', *Brain and Language*, vol.8, pp.133–61.

Caramazza, A. (1986) 'On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: the case for single-patient studies', *Brain and Cognition*, vol.5, pp.41–66.

Caramazza, A. and Badecker, W. (1989) 'Patient classification in neuropsychological research', *Brain and Cognition*, vol.10, pp.256–95.

Carpenter, K., Berti, A., Oxbury, S., Molyneux, A.J., Bisiach, E. and Oxbury, J.M. (1995) 'Awareness of and memory for arm weakness during intracarotid sodium amytal testing', *Brain*, vol.118, pp.243–51.

Charter, N. and Ganis, G. (1991) 'Double dissociation and isolable cognitive processes', in *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, Hillsdale, NJ, Erlbaum.

Chomsky, N. (1980) *Rules and Representations*, New York, Columbia University Press.

Cocchini, G. and Della Sala, S. (2010) 'Assessing anosognosia for motor and language impairments', in Prigatano, G. (ed.) *Study of Anosognosia*, Oxford, Oxford University Press.

Cocchini, G., Beschin, N., Cameron, A., Foutopolou, A., and Della Sala, S. (2009) 'Anosognosia for motor impairment following left-brain damage', *Neuropsychology*, vol.23, no.2, pp.223–30.

Cocchini, G., Beschin, N., and Della Sala, S. (2002) 'Chronic anosognosia: a case report and theoretical account', *Neuropsychologia*, vol.40, no.12, pp.2030–8.

Coltheart, M. (1980) 'Deep dyslexia: a review of the syndrome', in Coltheart, M., Patterson, K., and Marshall, J.C. (eds.) *Deep Dyslexia*, London, Routledge & Kegan Paul.

Coltheart, M. (1986) 'Cognitive neuropsychology', in Posner, M. and Marin, O.S.M. (eds.) *Attention and Performance, XI*, Hillsdale, NJ, Lawrence Erlbaum Associates Inc.

Coltheart, M. (1999) 'Modularity and cognition', *Trends in Cognitive Science*, vol.3, pp.115–120.

Coltheart, M. (2001) 'Assumptions and methods in cognitive neuropsychology', in Rapp B. (ed.) *The Handbook of Cognitive Neuropsychology*, Hove, Psychology Press.

Coltheart, M., Patterson, K., and Marshall, J.C. (1980) *Deep Dyslexia*, London, Routledge & Kegan Paul.

Cubelli, R., Marchetti, C., Boscolo, G., and Della Sala, S. (2000) 'Cognition in action: testing a model of limb apraxia', *Brain and Cognition*, vol.44, pp.144–65.

Davies, M., Davies, A.A., and Coltheart, M. (2005) 'Anosognosia and the two-factor theory of delusions', *Mind & Language*, vol.20, pp.209–36.

Della Sala, S. and Bates, P. (1996) 'Introduction. Research in cognition: the gateway to a sound rehabilitation', in Della Sala, S., Marchetti, C., and Turnbull, O.H. (eds.) *An Interdisciplinary Approach to the Rehabilitation of the Neurological Patient: A cognitive perspective*, Pavia, Fondazione S. Maugeri Edizioni PI-ME Press.

Durkin, M.W., Meador, K.J., Nichols, M.E., Lee, G.P., and Loring, D.W. (1994) 'Anosognosia and the intracarotid amobarbital procedure (Wada Test)', *Neurology*, vol.44, pp.978–9.

Dywan, C., McGlone, J., and Fox, A. (1995) 'Do intracarotid barbiturate injections offer a way to investigate hemispheric models of anosognosia?' *Journal of Clinical and Experimental Neuropsychology*, vol.17, pp.431–8.

Ellis, A.W. (1987) 'Intimations of modularity, or the modularity of mind: doing cognitive neuropsychology without syndromes', in Coltheart, M., Sartori, G., and Job, R. (eds.) *The Cognitive Neuropsychology of Language*, London, Lawrence Erlbaum.

Ellis, A.W. and Young, A.W. (1988) *Human Cognitive Neuropsychology*, Hove, Lawrence Erlbaum Associates.

Ellis, A.X., Della Sala, S., and Logie, R.H. (1996) 'The bailiwick of visuo-spatial working memory: evidence from unilateral spatial neglect', *Cognitive Brain Research*, vol.3, pp.71–8.

Fodor, J.A. (1983) *The Modularity of Mind*, Cambridge, MIT Press.

Fodor, J.A. (2000) *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*, Cambridge, MIT Press.

Gazzaniga, M.S., Ivry, R.B., and Mangun, G.R. (2002) *Cognitive Neuroscience: The Biology of Mind*, New York, W.W. Norton & Company.

Gerstmann, J. (1930) 'Zür symptomatologie der hirnlaesionen im ubergangsgebiet der unteren parietal- und mittelerem occipitalwindung', *Nervenarzt*, vol.3, pp.691–5.

Geshwind, N. (1965) 'Disconnexion syndromes in animals and man', *Brain*, vol.88, pp.237–94.

Gilmore, R.L., Heilman, K.M., Schmidt, R.P., Fennell, E.M., and Quisling, R. (1992) 'Anosognosia during Wada testing', *Neurology*, vol.42, pp.925–7.

Harlow, J.M. (1848) 'Recovery after severe injury to the head', *Publication of the Massachusetts Medical Society*, vol.2, pp.327–46.

Heilman, K.M., Rothi, L.J., and Valenstein, E. (1982) 'Two forms of ideomotor apraxia', *Neurology*, vol.32, pp.342–6.

Humphreys, G.W., Heinke, D.G., and Yoon, E.Y. (2006) 'Cognitive neuropsychology and computational modelling: the contribution of computational neuroscience to understanding the mind' in Senior, C., Russell, T., and Gazzaniga, M.S. (eds.) *Methods in Mind*, Cambridge, MIT Press.

Job, R. and Sartori, G. (1984) 'Morphological decomposition: evidence from crossed phonological dyslexia', *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, vol.36A, no.3, pp.435–58.

Kaplan, R.F., Meadows, M.E., Cohen, R.A., Bromfield, E.B., and Ehrenberg, B.L. (1993) 'Awareness of deficit after the sodium amobarbital (Wada) test', *Journal of Clinical Experimental Neuropsychology*, vol.15, p.383.

Levine, D.N., Calvanio, R., and Popovics, A. (1982) 'Language in the absence of inner speech', *Neuropsychologia*, vol.20, pp.391–409.

Lichtheim, L. (1885) 'On aphasia', *Brain*, vol.7, pp.433–484.

Liepmann, H. (1900) 'Das krankheitshildder der apraxie (Motorischen/Asymbolie)', *Monatschrift für Psychiatrie und Neurologie*, vol.8, pp.15–44, 102–132, 182–197.

Lu, L.H., Barrett, A.M., Schwartz, R.L., Cibula, J.E., Gilmore, R.L., Uthman, B.M., and Heilman, K.M. (1997) 'Anosognosia and confabulation during the Wada test', *Neurology*, vol.49, pp.1316–22.

Luria, A.R. (1966) *Higher Cortical Functions in Man*, London, Tavistock.

Luria, A.R. (1973) *The Working Brain*, London, Penguin.

Marie, P. (1906) 'Révision de la question de l'aphasie: la troisième convolution frontale gauche ne joue aucun role speciale dans la fonction du language', *Semaine Medicale*, vol.21, pp.241–7.

Marin, O.S.M., Saffran, E.M., and Schwartz, D.F. (1976) 'Dissociations of language in aphasia: implications for normal functions', *Annals of the New York Academy of Sciences*, vol.280, pp.868–84.

Marr, D. (1976) 'Early processing of visual information', *Philosophical Transactions of the Royal Society of London B*, vol.275, pp.483–524.

Marshall, J.C. and Halligan, P.W. (1988) 'Blindsight in visuo-spatial neglect', *Nature*, vol.336, pp.766–7.

Marshall, J.C. and Newcombe, F. (1966) 'Syntactic and semantic errors in paralexia', *Neuropsychologia*, vol.4, pp.169–76.

Marshall, J.C. and Newcombe, F. (1973) 'Patterns of paralexia: a psycholinguistic approach', *Journal of Psycholinguistic Research*, vol.2, pp.175–99.

McCloskey, M. (2001) 'The future of cognitive neuropsychology', in Rapp, B. (ed.) *The Handbook of Cognitive Neuropsychology*, Hove, Psychology Press.

Milner, B. (1966) 'Amnesia following operations on the temporal lobes', in Whitty, C.W.M. and Zangwill, O.L. (eds.) *Amnesia*, London, Butterworth.

Milner, B., Corkin, S., and Teuber, H-L. (1968) 'Further analysis of the hippocampal amnesia syndrome: 14 year follow-up study of H.M.', *Neuropsychologia*, vol.6, pp.215–34.

Morton, J. and Patterson, K. (1980) 'A new attempt at an interpretation, or, an attempt at a new interpretation', in Coltheart, M., Patterson, K., and Marshall, J.C. (eds.) *Deep Dyslexia*, London, Routledge & Kegan Paul.

Patterson, K., Marshall, J.C., and Coltheart, M. (1985) *Surface Dyslexia*, London, Lawrence Erlbaum.

Perret, E. (1974) 'The left frontal lobe in man and the suppression of habitual responses in verbal categorical behaviour', *Neuropsychologia*, vol.12, pp.323–30.

Poeck, K. and Orgass, B. (1966) 'Gerstmann's syndrome and aphasia', *Cortex*, vol.2, pp.421–7.

Popper, K.R. (1959) *The Logic of Scientific Discovery*, London, Hutchinson.

Rafal, R., Smith, J., Krantz, J., Cohen, A., and Brennan, C. (1990) 'Extrageniculate vision in hemianoptic humans: saccade inhibition by signals in the blind field', *Science*, vol.250, pp.118–21.

Ramachandran, V.S., Rogers-Ramachandran, D.C., and Cobbs, S. (1995) 'Touching the phantom limb', *Nature*, vol.377, pp.489–90.

Robertson, I.H., Halligan, P.W., Bergego, C., Hömberg, V., Pizzamiglio, L., Weber, E., *et al.* (1994) 'Right neglect following right hemisphere damage?' *Cortex*, vol.30, pp.199–213.

Robertson, L.C., Knight, R.T., Rafal, R., and Shimamura, A.P. (1993) 'Cognitive neuropsychology is more than single-case studies', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.19, pp.710–17.

Schnider, A. and Ptak, R. (1999) 'Spontaneous confabulators fail to suppress currently irrelevant memory traces', *Nature Neuroscience*, vol.2, pp.677–81.

Schnider, A., Ptak, R., von Däniken, C., and Remonda, L. (2000) 'Recovery from spontaneous confabulations parallels recovery of temporal conclusion memory', *Neurology*, vol.55, pp.74–83.

Schnider, A., von Däniken, C., and Gutbrod, K. (1996) 'The mechanisms of spontaneous and provoked confabulations', *Brain*, vol.119, pp.1365–75.

Scoville, W.B. and Milner, B. (1957) 'Loss of recent memory after bilateral hippocampal lesions', *Journal of Neurology, Neurosurgery and Psychiatry*, vol.20, pp.11–21.

Selnes, O.A. (2001) 'A historical overview of contributions from the study of deficits', in Rapp, B. (ed.) *The Handbook of Cognitive Neuropsychology*, Hove, Psychology Press.

Shallice, T. (1988) *From Neuropsychology to Mental Structure*, Cambridge, Cambridge University Press.

Sokol, S.M., McCloskey, M., Cohen, N.J., and Aliminosa, D. (1991) 'Cognitive representations and processes in arithmetic: inferences from the performance of brain-damaged subjects, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.17, pp.355–76.

Sperry, R.W., Gazzaniga, M.S., and Bogen, J.E. (1969) 'Interhemispheric relationships: the neocortical commissures; syndromes of hemisphere disconnection', in Vinken, P.J. and Bruyn, G.W. (eds.) *Handbook of Clinical Neurology, Volume 4*, Amsterdam, North-Holland Publishing Company.

Vallar, G. (1999) 'Methodological foundations of neuropsychology', in Denes, G. and Pizzamiglio, L. (eds.) *Handbook of*

Clinical and Experimental Neuropsychology, Hove, Psychology Press.

Vallar, G., Rusconi, M.L., and Bisiach, E. (1994) 'Awareness of contralesional information in unilateral neglect: effects of verbal cueing, tracing, and vestibular stimulation', in Umilta, C. and Moscovitch, M. (eds.) *Attention and Performance XV*, Cambridge, MIT Press.

Vuilleumier, P., Hester, D., Assal, G., and Regli, F. (1996) 'Unilateral spatial neglect recovery after sequential strokes', *Neurology*, vol.19, pp.184–9.

Ward, J. (2006) *The Student's Guide to Cognitive Neuroscience*, Hove, Psychology Press.

Warrington, E.K. and Shallice, T. (1969) 'The selective impairment of auditory verbal short-term memory', *Brain*, vol.92, pp.885–96.

Weiskrantz, L. (1986) *Blindsight: A Case Study and Implications*, Oxford, Oxford University Press.

Weiskrantz, L., Warrington, E.K., Sanders, M.D., and Marshall, J. (1974) 'Visual capacity in the hemianopic field following a restricted occipital ablation', *Brain*, vol.97, pp.709–28.

Wernicke, C. (1874) *Der Aphasische Symptomencomplex: Eine Psychologische Studie auf Anatomischer Basis*, Breslau, Germany, Cohn & Weigert.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

CHAPTER 14

NEUROIMAGING

Ingrid S. Johnsrude and Olaf Hauk

1 INTRODUCTION

Fifty years ago, almost the only way to relate human behaviour and cognition to the function of parts of the brain was to look for specific deficits in patients with damage confined to a particular region of the brain (acquired through, for example, stroke, injury, or surgery for neurological disease). Identifying the location of brain damage would often have to wait until after the patient's death, since the only way to examine brain structure in living people was the **pneumoencephalogram**, a low-resolution, poor-contrast x-ray that was neither painless nor harmless.

Although neuropsychological studies of brain-damaged people (see Chapter 13) remain extremely valuable, the development of functional imaging techniques has had an enormous impact on the way we assess human cognition. Detailed anatomical images, acquired through techniques such as magnetic resonance imaging (MRI), can now be used to identify structures in normal individuals (as shown in Figure 14.1) and localize damage in patients. Furthermore, it is possible to explore normal and abnormal brain function by combining detailed anatomical images from normal volunteers and patients with functional information from functional magnetic resonance imaging (fMRI), electroencephalography (EEG) and magnetoencephalography (MEG), transcranial magnetic stimulation (TMS), or, less commonly, from positron emission tomography (PET), single photon emission computerized tomography (SPECT), or near-infrared optical spectroscopy (NIRS).

Most neuroimaging techniques are easy and painless for volunteers. MEG (shown in Figure 14.2(a)) involves sitting in a device that looks rather like an old-fashioned beauty-salon hairdryer. Inside the headpiece are sensitive devices for measuring the magnetic fields produced by the brain (see Section 2.1.2). EEG (Figure 14.2(b)) involves putting electrodes, sensitive to the aggregate activity of large numbers of neurons, on the scalp with a conductive gel. Often, as is shown here, the electrodes are incorporated into a cap, which makes setting up the electrode array less time-consuming. In fMRI (Figure 14.2(c) and (d)), as in PET (e), the volunteer is lying comfortably on a bed that slides into the scanner. NIRS (f) measures changes in the brain's oxygen absorption based on optical properties of haemoglobin (blood), using light in the near-infrared range (700–1000 nm in wavelength). As with EEG, the detector array is a lightweight net that sits on the scalp, so that cortical activity can be assessed while participants are comfortably seated in an ordinary chair (f), or on a parent's lap (NIRS is often used with infants).

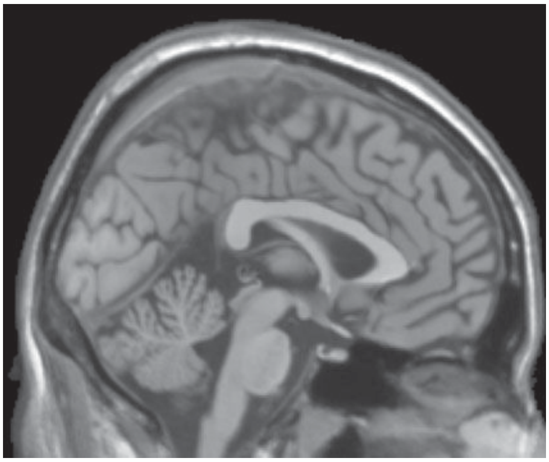


FIGURE 14.1 A magnetic resonance image of the brain of a neurologically normal individual. With fine spatial resolution, different brain structures can be easily identified: here, the image reveals different tissues such as grey matter (dark grey), white matter (light grey), fat (white), and fluid (black). Similarly, anatomical abnormalities can be identified, delineated, and localized with a high degree of accuracy.

Transcranial magnetic stimulation (TMS) is also used while volunteers are comfortably seated (g). It can safely and temporarily disrupt function in a specific brain region, essentially producing a temporary 'lesion', in order to evaluate the importance of that area of the brain to performance on a given task. All of these tools are powerful additions to the traditional neuropsychological method of cataloguing deficits in patients (see Chapter 13) in order to understand the relationship between brain and behaviour.

Each neuroimaging technique has its own advantages and disadvantages, many of which come under the general heading of **resolution**. The resolution of a technique indicates how far away, in space or time, two events have to be before they can be distinguished as two separate events. For example, EEG and MEG, which measure the electrical (EEG) or magnetic (MEG) fields generated by the activity of groups of neurons, have very good **temporal resolution** (in the order of milliseconds). This means that two events happening very quickly one after the other can be distinguished. However, as will be explained in the next section, the **spatial resolution** of EEG and MEG is only in the range of several centimetres, and the techniques are not equally sensitive over the whole brain. This means that EEG and MEG data cannot be linked to a particular region of the brain with much confidence. fMRI, NIRS, and often PET are sensitive to signals that depend on changes in the brain's blood supply: specifically, they measure brain metabolism (of oxygen in particular). Although fMRI can be used to localize active tissue with great spatial precision, the temporal resolution of these techniques is quite poor (a few hundred milliseconds at best for the most sophisticated experimental fMRI designs) (see Figure 14.3 and also Table 14.1 on page 465).

These tools each have their own special place in answering questions about cognition. If a question concerns 'where' in the brain increased activity is observed during perception of a stimulus or performance of a task, then fMRI (or sometimes PET or NIRS) would be helpful. fMRI studies, for example, have shown that different parts of the ventral visual processing pathway in the temporal–occipital cortex are specialized for processing different kinds of things such as faces (and other parts of the body), places, and tools. If, however, the question concerns 'when' a process occurs or the order in which different brain areas (identified perhaps using fMRI) become active, then EEG and MEG can be helpful. For example, both EEG and MEG can be used to show that, when a word is heard, frontal and temporal areas known to be important for language become active *after* auditory cortex does. Unlike the other methods, both TMS and the lesion method (the classic neuropsychological method of assessing the cognitive function of patients with focal brain lesions; see Chapter 13) reveal whether a particular brain area is *necessary* for a particular function, not just whether it is *involved* when a task tapping that function is performed.

This chapter introduces these various techniques, the assumptions underlying the methods, how they work, and what their advantages, disadvantages, and limitations are. We review how experiments using these techniques are designed to ask cognitive and neuroscientific questions, how data are analysed, and how results are interpreted. We begin by presenting methods based on measuring the brain's electrical activity (EEG/MEG), then present methods based on brain metabolism (fMRI and NIRS, but also PET and

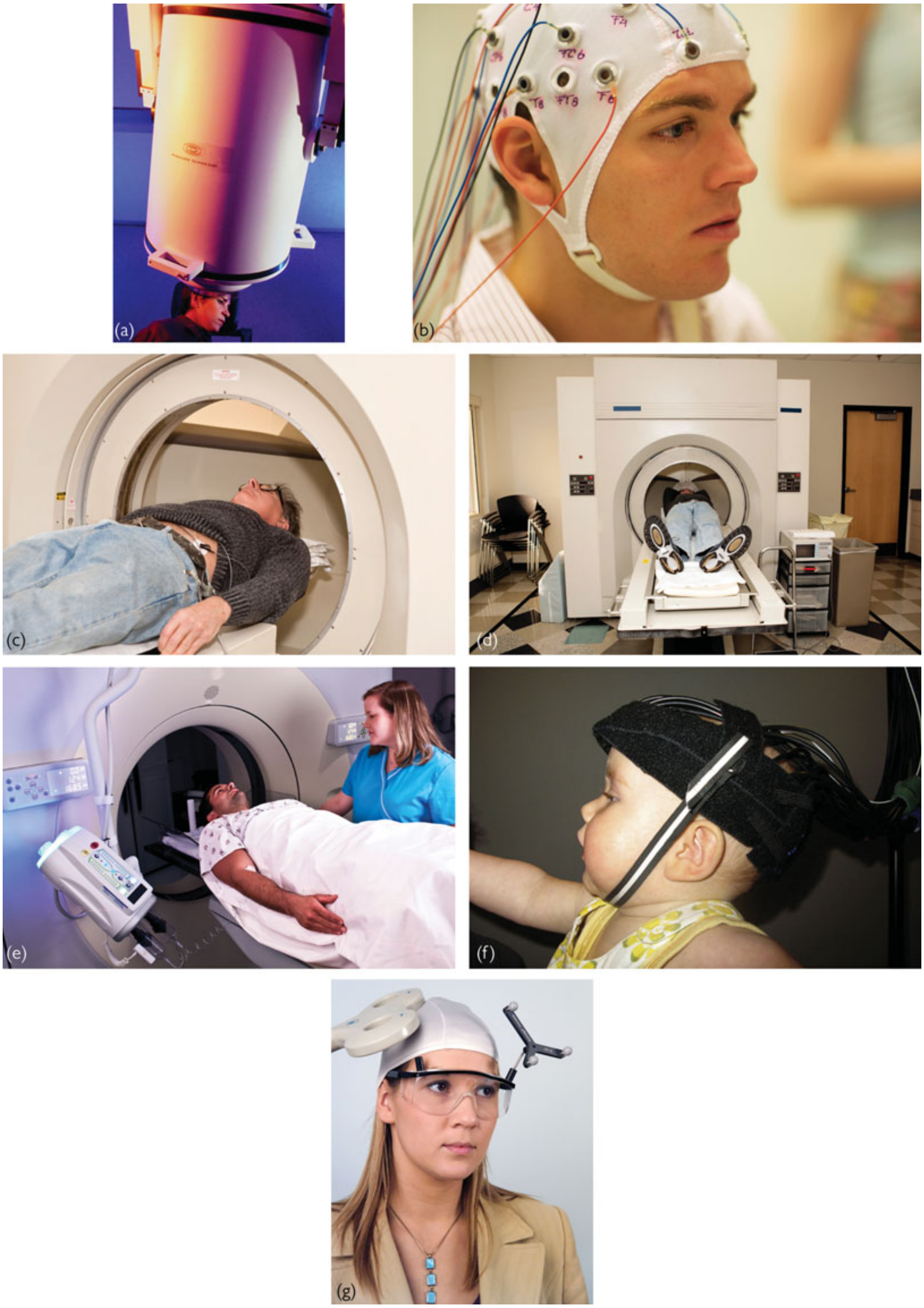
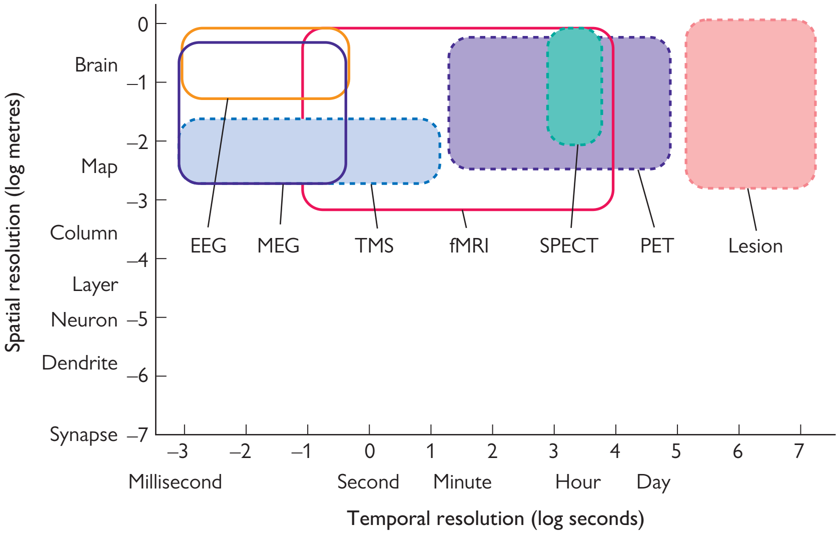


FIGURE 14.2 (a) MEG, (b) EEG, (c and d) fMRI, (e) activation PET, (f) NIRS, (g) TMS.

SPECT), and then briefly present transcranial magnetic stimulation (TMS) as a tool for inducing 'reversible lesions'. The information presented here should help you to understand papers on topics in cognitive psychology or cognitive neuroscience. It may also help you to begin to think creatively about how particular questions regarding the relationship between brain and behaviour might best be addressed.



KEY:

EEG:	electroencephalography	MEG:	magnetoencephalography
fMRI:	functional magnetic resonance imaging	TMS:	transcranial magnetic stimulation
PET:	positron emission tomography	SPECT:	single photon emission computed tomography
Lesion:	behavioural assessment of patients with focal brain lesions		

FIGURE 14.3 The ranges of temporal and spatial resolution of methods used to study human brain function, based on performance with typical hardware and commonly used experimental protocols. Shading indicates invasive techniques. The vertical axis includes references to 'maps' (e.g. visual or motor maps of space and the body); 'columns' (which are functionally specialized regions within a particular brain region – e.g. orientation columns in the visual cortex contain neurons that all fire when viewing a line that is angled a particular way); and 'layers' (the cortex of the brain is composed of six layers). *Source: adapted from Cohen and Bookheimer, 1994; Churchland and Sejnowski, 1988*

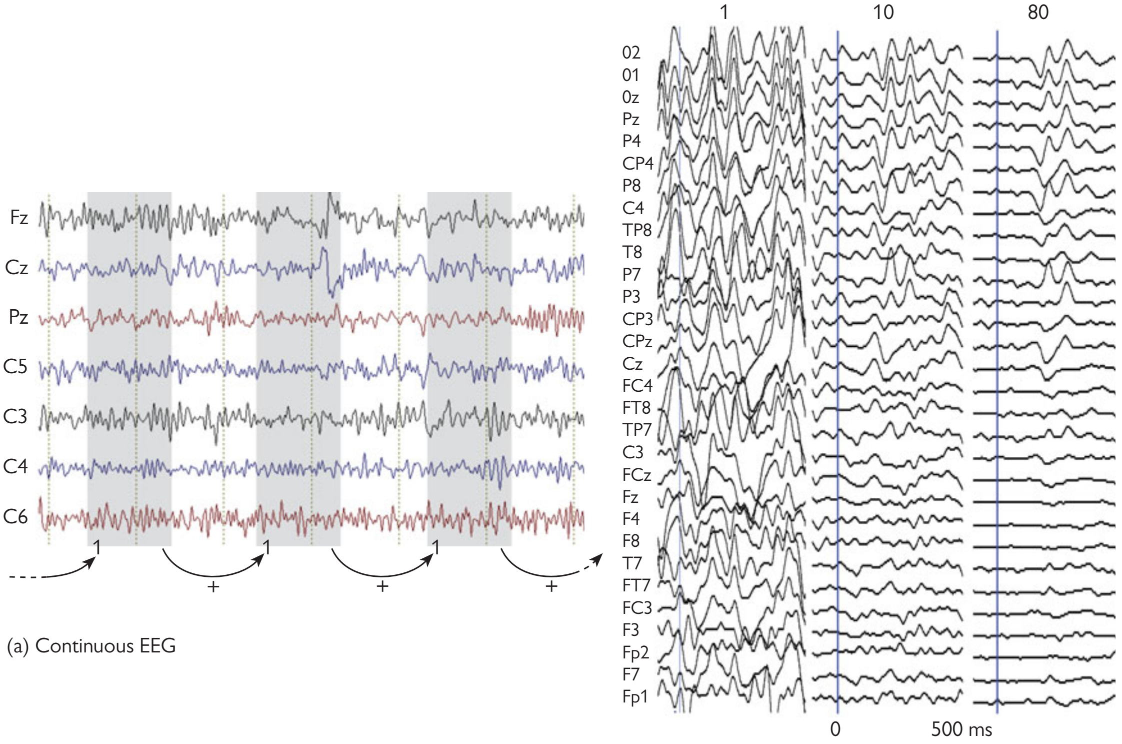
2 EEG AND MEG: MEASURING THE TIMING OF ELECTRICAL ACTIVITY IN THE BRAIN

The electrical activity of nerve cells (neurons) in the brain produces electrical currents spreading through the head. These currents, representing the integrated activity of many thousands of neurons, also reach the scalp surface. The resulting voltage differences on the scalp can be recorded as the electroencephalogram (EEG). The currents inside the head also produce magnetic fields, which can similarly be measured above the scalp surface as the magnetoencephalogram (MEG). EEG and MEG reflect brain electrical activity with millisecond (1/1000th of a second) temporal resolution, and provide the most direct correlate of 'on-line' brain processing that can be obtained non-invasively (without physically entering the body).

2.1 Collecting data

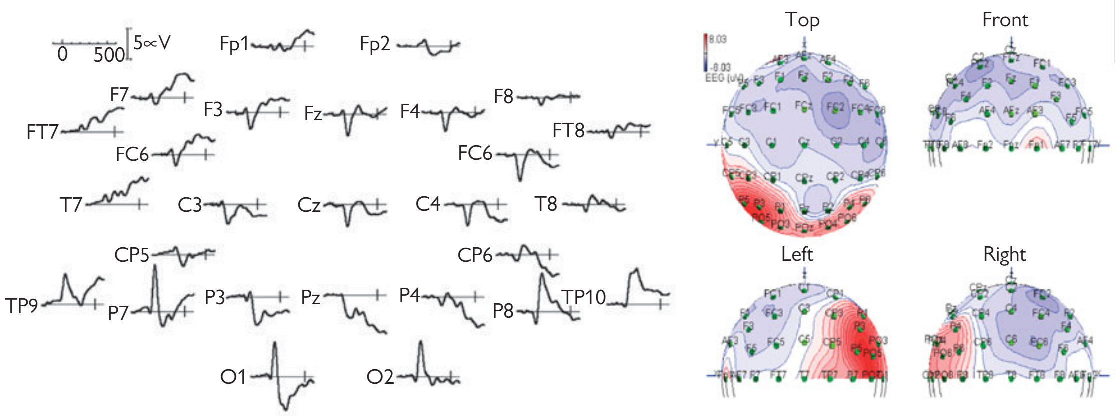
2.1.1 EEG recording

The continuous or spontaneous EEG (i.e. reflecting continuous voltage fluctuations as a function of



(a) Continuous EEG

(b) Averaged EEG



(c) Time plot of ERP ('ERP curves')

(d) Topography of ERP at 90 ms

FIGURE 14.4 Data pre-processing of event-related EEG and MEG data: (a) Illustration of the average procedure used in standard event-related potential (ERP) analysis (see Section 2.2.1). (b) Averages consisting of 1, 10, and 80 time segments, respectively. Note that the pre-stimulus interval ('baseline') flattens with increasing number of segments, while the event-related response becomes less distorted by noise, i.e. the 'signal-to-noise ratio' (SNR) increases. (c) The averaged curves can be plotted for each electrode or sensor location separately. Channels are plotted according to their relative position on the lead. (d) A continuous topographic plot over the head. Red colour indicates positive and blue colour negative signals. The picture shows the event-related potential 90 ms after visual presentation of a word, with two distinct positive peaks over left and right occipital scalp areas, presumably reflecting processing of visual stimulus features.

time, that can be viewed directly for each electrode during the recording (see Figure 14.4(a)) can be clinically very helpful. For example, it can be used to diagnose epilepsy or another seizure disorder, predict seizures, detect abnormal brain states, or classify sleep stages.

However, the amplitude of the continuous EEG signal is of the order of several tens of microvolts (μV, or one millionth of a volt), i.e. about a million times weaker than that of household batteries. It therefore requires special amplifiers that amplify the signal and convert it to a digital signal that can be stored and processed on a computer. Recordings are taken virtually simultaneously at all recording sites (which can range in number from several tens to several hundreds) and at different time points (usually every few milliseconds). The time difference between sampling points is called the **sampling interval**, and the **sampling rate** is given by the number of samples per second (unit Hertz), i.e.

$$\frac{1000\,\text{ms}}{(\text{sampling interval in ms})}$$

In typical EEG or MEG studies, a sampling rate of 200–500 Hertz (Hz) is common (corresponding to rates between 200 and 500 samples per second, or to sampling intervals of between 5 and 2 ms). EEG is recorded using a system of **electrodes** placed on the scalp. Electrodes are usually little metal discs that are pasted onto the scalp among the hair. They are either mounted on a cap (e.g. if a large number of electrodes are used), or can be fixed one by one. To assure good data quality, a conductive substance usually has to be inserted between the electrodes and the scalp to improve the contact between the two. A particular advantage of EEG compared to MEG, fMRI, and PET is that once the electrodes are fitted the participant is allowed to move their head within certain limits. This makes this technique suitable for infants or for patients where head movement cannot be restricted.

To compare results across different studies, electrode locations must be standardized and locations need to be easily determined for individual subjects. Common is the so-called **extended 10/20 system** (see Figure 14.5), where electrode locations are defined with respect to fractions of the distance between **nasion** and **inion** (i.e. root of the nose to 'the little bump on the back of your head', or 'front-back') and the **pre-auricular points** (the little skin lobules in front of your ear canals, or 'left-right', also known as left pre-auricular or LPA and right pre-auricular or RPA). Electrode labels like 'FPz' or 'PO4' indicate the area of the head ('Fronto-Polar', 'Parieto-Occipital'), odd numbers mark left-hemispheric and even numbers mark right-hemispheric locations, and the lower the number, the nearer the electrode is to the mid-line ('z' indicates a location on the mid-line). 'TP8' ('Temporo-Parietal') is therefore more to the right than 'TP2', and 'F5' ('Frontal') is more to the left than 'F1'.

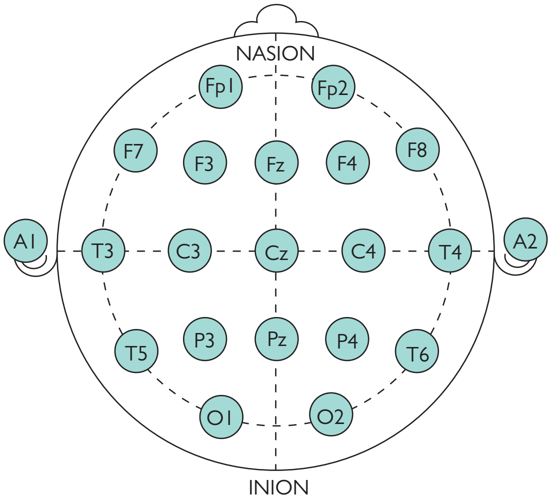


FIGURE 14.5 The extended 10/20 system.

Electrical voltages (also called potentials) are only defined with respect to a **reference**, i.e. an arbitrarily chosen 'zero level'. The choice of the reference may differ depending on the purpose of the recording. This is similar to measures of height, where the zero level can be at sea-level if we refer to the height of mountains, or at ground level if we refer to the height of a building, or floor level for the height of a person.

For each EEG recording, a **recording reference** has to be selected in advance. Ideally, this reference would be affected by global voltage changes (e.g. slow voltage shifts due to perspiration) in the same manner as all the other electrodes, such that brain non-specific activity is subtracted out by the referencing. Also, the reference should not pick up signals that are not intended to be recorded, like heart activity, which would be 'subtracted in' by the referencing. Possible locations for reference electrodes are the ear lobes, the tip of the nose, or the mastoids (i.e. the bones behind the ears). It is possible to 're-reference' the data after the recording, or 'off-line'. A common choice for multi-channel recordings (e.g. more than 32 channels) is the average reference: the average across all electrodes is subtracted from each individual electrode at each time sample.

Source estimation involves trying to infer, on the basis of the pattern of electrical activity at the scalp surface, where the critical processing occurs inside the brain (see Section 2.3). To do this using EEG data, one needs to know the positions of the electrodes. As a rough approximation, one can use the standardized electrode positions associated with electrode labels (e.g. 'PO7'). If more accurate localization is required (e.g. for co-registration with MRI images), electrode positions have to be measured using special 'digitizers'.

2.1.2 MEG recordings

The magnetic field, in contrast to the electrical voltage or potential, has a direction, usually illustrated by magnetic field lines. The strength of the magnetic fields produced by the human brain is less than a millionth of the Earth's magnetic field. Such small field strengths can be measured by so-called superconducting quantum interference devices (SQUIDs). Because MEG sensors are so very sensitive, they have to be well-shielded against magnetic noise from the outside. The shielding of the laboratory and the SQUID technology make an MEG device about 100 times more expensive than an EEG system, and the required permanent helium-cooling (required for the superconductivity of the detectors) imposes considerable maintenance costs.

Because MEG sensors have to be cooled by liquid helium, they are permanently mounted within a helmet device, several centimetres from the scalp surface. The latest 'whole-head' systems have around 300 sensors. For further data processing, in particular for source estimation (see Section 2.3), it is crucial to know the relative position of the sensors with respect to the head. Therefore, small coils are attached to the participants' heads, which can be localized within the MEG system relative to the sensors. Landmarks on the participants' heads (e.g. nasion, LPA, RPA; or a distribution of points across the scalp) are also measured in relation to these coils. These landmarks can later be aligned to the participants' structural MRI images, for example. Via the position of the coils, it is then also possible to co-register the locations of the MEG sensors to the MRI image.

2.1.3 Basic differences between EEG and MEG

Both EEG and MEG are sensitive to the topography of the brain – the details of the structure of its surface. The brain surface is quite dramatically folded, and whether the signal arises in the convex tops of the folds (gyri) or the concave depths of the folds (sulci) affects sensitivity to neuronal current sources (known as dipoles). MEG is particularly insensitive to radial dipoles, those directed towards or away from the scalp (like at the top of a gyrus). It mainly 'sees' tangential dipoles, which are parallel to the scalp. The reason for this is that if the head were a perfect sphere (a reasonable first approximation for an average head shape), all magnetic fields resulting from a radial dipole (including the volume currents) would cancel each other out at the surface (a more detailed explanation would involve a fair amount of physics). EEG is sensitive to both radial and tangential dipoles.

A special case of a radial dipole is a dipole located at the centre of a sphere: it must necessarily point away from the centre, and is therefore radial and 'invisible' for MEG. Roughly speaking, the nearer a source is to the centre, the 'more radial' it is. This implies that the deeper a source is in the head, the less visible it is for MEG. This seems to be a disadvantage for MEG, but turns into an advantage if superficial and tangential sources (e.g. in sulci of the cortex) are targeted, since in that case the corresponding MEG signal is less contaminated than EEG by other possibly disturbing sources. Furthermore, it has been shown that superficial and tangential sources can be localized with more accuracy than with EEG, particularly in somatosensory and some auditory parts of the brain. MEG has therefore been widely used to study early acoustic and phonetic processing, and the plasticity of the somatosensory and auditory systems.

The basics of EEG were invented in the 1920s, mainly with medical applications in mind. EEG is widely used in clinical diagnostics and research. The first MEG systems (starting with 1-channel!) appeared in the late 1960s, and the first commercial systems took off in the 1980s. The development was mainly driven by engineers and physicists. This, and possibly the fact that MEG is much more expensive than EEG, often makes these two look like very different methods with very different applications. It is important to keep in mind, however, that both techniques measure the same type of brain activity from slightly different angles. As a general rule, EEG 'sees more' of the brain, but is less able to localize this activity. MEG 'sees less' of the brain, but what it sees it can localize better. As we will see, a combination of EEG and MEG is superior for localization than either method on its own. If localization is not the main focus of a study, EEG may often be sufficient, for example in many clinical environments.

2.2 Basic analysis of EEG and MEG data

2.2.1 Computing event-related potentials (ERPs) and event-related fields (ERFs)

Whereas the continuous EEG can be useful clinically, the use of EEG and MEG to investigate more specific perceptual or cognitive processes requires more sophisticated data processing. Usually, the electrical signals resulting from a particular cognitive event (e.g. hearing a word) are averaged over many trials (see Figure 14.4 (b)): a data epoch of a specified length (e.g. starting 100 ms before word onset, and ending 1 s after) is cut out of the data for each stimulus, and then all of these epochs are averaged together. During averaging, every epoch is checked for artefacts (see below), e.g. whether the amplitude of the signal is within a certain allowed range. Epochs that are likely to contain an artefact (i.e. a large signal not originating from the brain) should not be included in the average.

This method assumes that the brain response does not markedly change its timing or spatial distribution during the sequence of trials. This averaging procedure can be applied to all stimuli present in the study, or for subgroups of stimuli that are to be compared to each other. During averaging, any random fluctuations across epochs should cancel each other out, since they might have positive amplitudes in one segment, but negative amplitudes in another. In contrast, any brain response that is time-locked to the presentation of the stimulus will add up constructively and be reflected in the average activity across trials. These averaged data are usually referred to as the event-related potential (ERP) in the case of EEG, and event-related field (ERF) in the case of MEG. These have characteristic shapes in time (Figure 14.4 (c)) and space (Figure 14.4 (d)), which can be analysed into a series of prototypical deflections (often referred to as 'components') in their time course. The signal at any given time point can be interpolated between electrodes or sensors to yield a continuous distribution, or topography, over the whole head. Some deflections that exhibit characteristic time courses and topographies for a particular stimulus type or paradigm (e.g. always present for auditory stimuli, or for auditory stimuli that are odd-ones-out in a sequence) are conventionally labelled according to their polarity at the peak – P(ositive) or N(egative) – and the typical latency of their occurrence. For example, 'auditory N100' refers to a negative potential at frontal electrode sites, usually observed 100 ms after the onset of an auditory stimulus (e.g. a simple tone, a speech sound, etc.).

The strength and pattern of these components can be measured, and their dependence on task requirements, on stimulus parameters, on degree of attention, and so on can be evaluated. Components known to be related to perceptual processing in the tactile, visual, and auditory domains (known as evoked potentials, or evoked magnetic fields) can be used by clinicians to check the connections between the peripheral and central nervous systems, much like an electrician might use a voltmeter to diagnose a faulty electrical cable.

For example, the presentation of simple tones is normally succeeded by specific components in the EEG signal. These are classified according to their time of occurrence into brain-stem responses (until approximately 20 ms after stimulus onset), middle latency responses (20–50 ms approx. after onset), and slow or late responses (later than 50 ms approx. after onset). It is known that the brain-stem responses originate from parts of the brain stem and/or from a cranial nerve (hence their name), while the late responses are cortical in origin. This information is used in objective hearing tests. Such measures are used to test the integrity of the auditory system from brain stem to cortex, and to screen for possible congenital deafness in newborns and infants, with whom conventional hearing tests that require a behavioural or verbal response to sound cannot be used. Finally, such measures can also be used in people who are suspected of faking hearing loss for insurance reasons. By detecting abnormalities in these components of the EEG signal, the stage at which damage to the auditory system has occurred can be determined.

2.2.2 Evoked and induced brain responses

The averaging procedure described above assumes that there is something like an 'average brain response', i.e. a signal with roughly the same time course in every epoch for a particular stimulus category. This means that the EEG response evoked by stimulation happens at the same time, so that the 'peaks' and 'troughs' of the evoked waves line up nicely (i.e. they are 'in phase', or 'phase-locked'). But what if the timing changes slightly from epoch to epoch? If this jitter across epochs is small compared to the time-scale of the signal, one does not have to worry. But what if the signal consists of a rapid sequence of positive and negative deflections (e.g. an 'oscillation')? Then the jitter may sometimes result in a positive value in some epochs and negative values in other epochs. Across epochs, this may average out to zero. This could be the case with so-called 'gamma oscillations' – a signal in the frequency range above 30 Hz (though definitions differ slightly), which has been associated with complex stimulus processing such as Gestalt perception. These gamma oscillations may occur time-locked to a stimulus (e.g. always between 200 and 300 ms after picture onset). However, the 'phase' – the exact occurrence in time of peaks and troughs – may vary across epochs. These signals are called 'induced', in order to distinguish them from the above-mentioned 'evoked' signals that are both time- and phase-locked to the stimulus. In order to extract induced signals from the data, the frequency power spectrum (i.e. the signal amplitude for different frequency bands) has to be computed for every epoch separately. The power spectrum only contains information about the amplitude of the signal for a certain frequency in a certain latency range, but no phase information. If there are consistent gamma oscillations between 200 and 300 ms after every picture, the spectral power in the gamma frequency range should be consistently above baseline level in this latency range across epochs. This method has been applied to studies on Gestalt perception, as illustrated in Figure 14.6. For example, it has been found that gamma band activity is usually larger when an incomplete but meaningful pattern is perceived, compared to either meaningless or easily recognizable patterns.

Participants were presented with visual stimuli that either did not evoke an association with a particular object (top left of Figure 14.6), or could be interpreted as edges of a triangle ('Kanizsa' triangle, bottom left). A stimulus can generate two types of brain response: evoked, i.e. a signal with the same time course after each presentation of the stimulus ('time-locked and phase-locked' to stimulus onset), such that across epochs the signals average together; and induced, i.e. a signal that is similar across different presentations of the stimulus but may move forward and backward in time a bit ('time-locked but not phase-locked'), such that inter-epoch jitter may result in cancellation of signals during the averaging process (illustrated at top right of Figure 14.6). These types of brain response can occur in different frequency ranges, i.e. for signals that change with different speed over time. At high frequencies, the change occurs very quickly, while at low frequencies it is slower. One can visualize the signal strength at a particular frequency against the time measured from stimulus onset in a so-called 'time-frequency plot'. The time-frequency plots in Figure 14.6 show the results for the meaningless stimulus configuration (top) and the Kanizsa triangle (bottom) for one electrode (Cz). A peak of 'gamma band activity' first occurred at about 100 ms and around 40 Hz

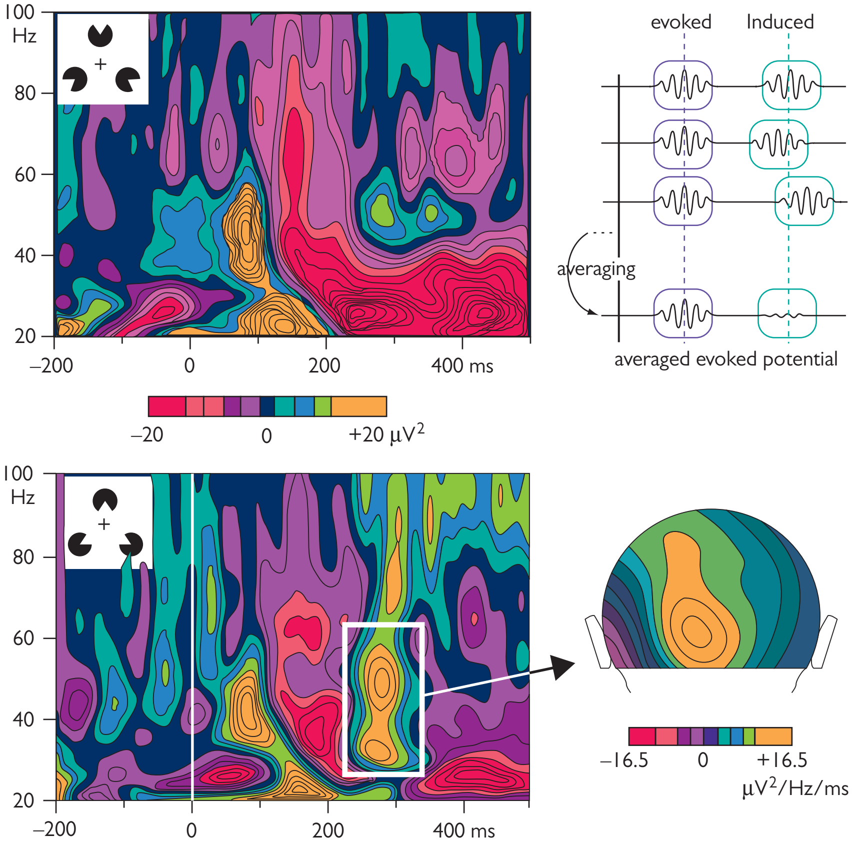


FIGURE 14.6 Evoked and induced oscillatory brain responses. *Source: Based on Tallon-Baudry and Bertrand, 1999, Figures 1 and 2*

(the so-called 'gamma' frequency range) for both stimulus types. A second peak, around 280 ms in a similar frequency range, was only present for the illusory triangle. Although evoked and induced activity were combined in this plot to save space, more detailed analysis revealed that the first peak represented evoked, and the second peak induced, brain activity. The evoked response is assumed to reflect basic early visual processing, while the induced activity is associated with spatial feature-binding processes required to perceive the illusory triangle. The topography of the induced activity (bottom right, back view) is consistent with neural generators in posterior parts of the brain, such as visual cortex.

2.2.3 Noise and artefacts

The measured signal not only contains the signal of interest, but also noise and artefacts. For example, during the recording EEG and MEG signals are undergoing slow shifts over time, such that the average level might differ considerably across channels, although the smaller and faster signal fluctuations are similar. These signal shifts can be due to slow brain activity, but can also be caused by perspiration (in the case of EEG), muscle tension, or other noise sources. Obviously, one should attempt to remove anything from the data that is not related to the processes of interest. The noise level in the averaged data decreases as the number of trials increases (specifically, noise decreases as a function of the square root of the number of trials). This poses a considerable limitation on the duration of an experiment, and should be carefully considered prior to setting up a study. A typical number of trials per experimental condition in EEG and MEG studies on cognitive processes is approximately 50–100, though in some cases this number can be lower or higher.

A small number of trials can be compensated for, in part, by recording data from more participants. In general, however, the rule is 'the more the better'.

Two other standard procedures for reducing noise are **filtering** and **baseline correction. Filtering** is applied to a data set to remove any frequencies that are not part of the signal of interest (i.e. the ERP/ERF components under investigation). For example, high frequency ranges (>40 Hz) are most prone to contamination by muscle-related noise activity, or line-noise from nearby electrical devices. Therefore, **low-pass filters** are used to attenuate frequencies above a certain **cut-off frequency** (i.e. they 'let pass' the lower frequencies). Analogously, high-pass filters can remove slow signal shifts. A combination of a low-pass and a high-pass filter is called a **band-pass filter**. A typical band-pass filter for ERP/ERF studies would be 0.1–30 Hz, but this depends on the signal of interest and the purpose of the study.

In most ERP and ERF studies, a 'baseline interval' is defined as several tens or hundreds of milliseconds before stimulus onset. For each recording channel, the mean signal over this interval is computed, and then subtracted from the signal at all time points. This procedure is usually referred to as **baseline correction**. It is crucial in ERP and ERF experiments to ensure that an observed effect (e.g. an amplitude difference between the components evoked by two sorts of stimuli) is not already present in the signal before the stimuli were presented, since such differences could not possibly reflect stimulus effects and must therefore be artefactual (i.e. a result due not to changes in the independent variable but perhaps to noise, or to the way equipment is set up).

A particularly annoying type of noise in both EEG and MEG recordings are **eye movements** and **muscle artefacts** (e.g. from face or neck muscles). Muscle artefacts can often be avoided by providing a comfortable recording environment for the participants, and by providing comprehensive instructions. They are usually reflected in higher frequency ranges (above 30 Hz), and may be reduced using low-pass filters.

Eye movements (and eye blinks in particular) are more difficult to control. In reading or visual search tasks, for example, eye movements may occur systematically with the stimulus or task. In many experiments, visual stimuli are therefore presented within a visual angle of less than 4 degrees (foveal presentation) and very briefly (less than 150 ms, 'tachistoscopically') in order to minimize systematic eye movements. Eye movements can be monitored using **EOG (electrooculogram)** electrodes placed around the eyes. Algorithms can be used to employ this information to reduce the amount of eye-related signal in the data. However, avoiding artefacts is always better than dealing with them afterwards. One should therefore always make the recording environment and experimental procedure as comfortable as possible (e.g. provide enough breaks and opportunity to practise).

2.3 Estimating neuronal sources: the inverse problem

The estimation of the sources in the brain that generate the electrical and magnetic activity recorded by EEG or MEG is usually called the **inverse problem**. What makes this a 'problem' is the fact that we are trying to identify the cause of the EEG or MEG signal from its effect, and unfortunately many different causes might lead to the same effect. This **inverse problem** is comparable to reconstructing an object from its shadow on the wall of your room. If you already know that there are only a few possible objects, then the choice might be easy. However, in some cases it may be very hard to know what is casting the shadow (e.g. if you just see a square shadow, is it a wardrobe or a bookshelf?). Furthermore, the more similar the shapes of the objects in question are, the more difficult it will be to identify which object causes the shadow.

In the case of EEG and MEG analysis, the shadow represents the data, or more precisely the distribution of the voltage or magnetic field across the sensors. What must the current distribution in the brain look like in order to produce the observed pattern? Unfortunately, there is no unique answer. Even if the

EEG and MEG signals were measured simultaneously at infinitely many points around the head, the information would still be insufficient to uniquely compute the distribution of currents within the brain that generated these signals (this is known as the Helmholtz principle, another expression for the inverse problem).

2.3.1 Approaches to deal with the inverse problem

All is not lost! Researchers can develop models of the underlying brain activity and test those against the data in order to reduce the number of plausible solutions, and figure out the most likely pattern of brain activity generating the observed pattern of EEG or MEG data. One can distinguish two broad strategies:

1. Approaches that make specific modelling assumptions, like the number of distinct activated brain regions and their approximate locations (for example, that there are two focal sources, one in the left and the other in the right auditory cortex, 80 ms after presentation of a tone).
2. Approaches that make as few assumptions as possible and focus on rough features of the current distribution (for example, that a language task will result in more left-hemispheric activation compared with a mental rotation task).

Examples of these two approaches, and how the models relate to the neuronal sources, are shown in Figure 14.7. In a simulation, two electric sources (dipoles) were placed, one in the left and one in the right hemisphere. The electric potential distribution produced by both dipoles together is shown as a contour plot in the left image in Figure 14.7(c). From mere visual inspection of the potential distribution alone, neither the location of the dipoles nor even the number of sources can be deduced. The right image shows a distributed source solution computed from the same data. It exhibits one peak of activation in each hemisphere, around the true dipole locations. Though the 'blurring' inherent to this method does not allow precise localization of the sources, it can provide information about their number (in our example two) and approximate location (here one source in each hemisphere at roughly symmetrical positions). The images on the left and right represent solutions to the same problem (i.e. the source models produce the same potential distribution over 148 recording channels).

If the number of activated brain regions that generate the recorded signal is assumed to be known, but their locations are still to be determined, then so-called dipole models can be applied (Figure 14.7(c), left). The locations and orientations of dipoles are systematically varied until the best fit between the predicted and the measured potential is achieved. If the modelling assumptions fit reality well, then the result can be very accurate (in the millimetre range for MEG and in the centimetre range for EEG). These methods are usually applied if perceptual processes and early-evoked components are studied (e.g. somatosensory, auditory, and visual processes up to 150–200 ms after stimulus onset).

If there is no information about the number of sources, methods with less restrictive modelling assumptions – so-called distributed source models – can be used. A large number of dipoles (hundreds or thousands) are distributed across the brain volume or the brain surface. Their strengths are then estimated such that their summed activity explains the measured signal and an additional criterion is fulfilled, e.g. that the overall source strength is kept minimal ('minimum norm criterion'). An example for such an estimate is shown on the right of Figure 14.7(c). At least a 'blurred' version of the real source distribution can often be obtained. Peaks in this estimated current distribution might correspond to centres of activity in the real current distribution. These methods are preferred if modelling assumptions for dipole models cannot be justified, such as in complex cognitive tasks or with noisy data.

No single approach is good for EEG and MEG source estimation. The choice of correct method requires insight into the modelling assumptions of the method, as well as experience.

2.3.2 EEG and MEG: localization issues

Although the signals of EEG and MEG are generated by the same sources (electrical currents in the brain),

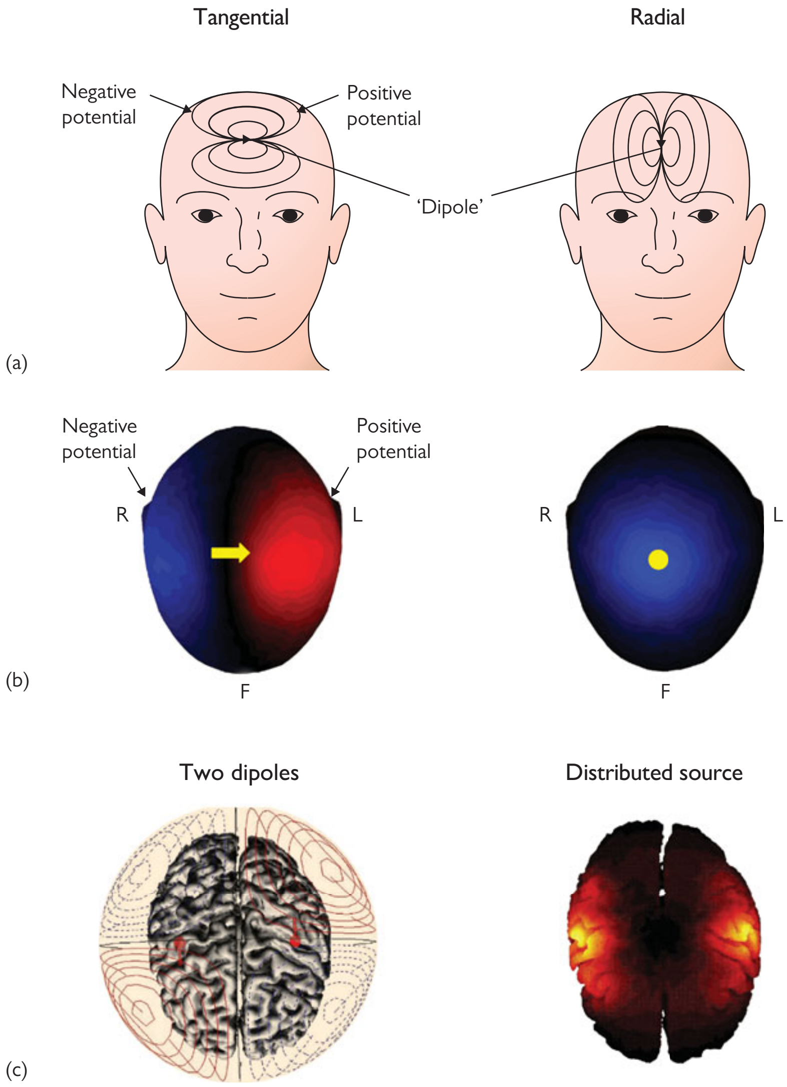


FIGURE 14.7 The generation and modelling of EEG/MEG signals: (a) The summed electrical activity within a small area of cortex can be described by a so-called 'dipole'. This dipole describes the strength and direction of the current flow within this area. Each dipole has a positive scalp potential in the direction in which it points, and a negative potential in the opposite direction. (b) (*left*) Potential distribution for a tangential dipole (pointing left to right), and (*right*) potential distribution for a radial dipole (pointing downwards). The magnetic field of a tangential dipole is roughly perpendicular to its potential distribution (not shown). (c) Illustration of the 'inverse problem' in EEG/MEG analysis: the images both represent different ways in which the same surface electrical and magnetic activity could be produced.

they are sensitive to different aspects of these sources (as described above). This could be compared to viewing the shadows of the same object from two different angles. It should not be surprising then that the information from EEG and MEG is complementary, and combining the two in source estimation usually yields more accurate results than using either method alone. However, this also means more work, and not everybody has both EEG and MEG available. It is therefore important to decide at the beginning of a study: 1) whether localization is required at all; 2) whether absolute anatomical localization is required (e.g. in combination with structural MRI images), or whether relative localization is sufficient (e.g. 'left more than right'); and 3) whether the sources are most visible in EEG, MEG, or both.

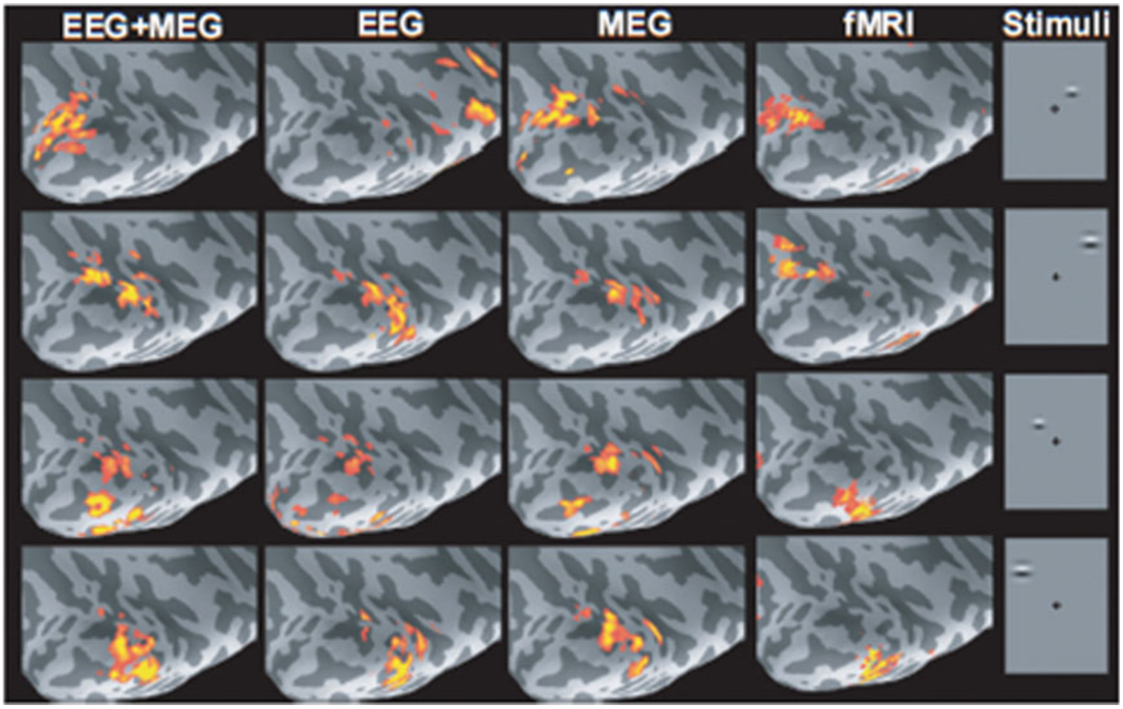


FIGURE 14.8 Comparison of source estimation among fMRI, EEG, and MEG. Simple visual stimuli were presented in different visual quadrants. Brain activity was localized using fMRI, MEG alone, EEG alone, and a combination of EEG and MEG. The images show the peaks (upper 1%) of activation distributions on the inflated ('flattened') surface in posterior cortex. In this experiment, the localization based on a combination of EEG and MEG most closely resembles the localization in fMRI. Note that this result was obtained using a very simple experimental paradigm, and a comparison (or combination) of fMRI with EEG/MEG may be much more complicated for cognitively more complex tasks. *Source: Based on Sharon et al., 2007, Figure 2*

While for EEG and MEG we can safely assume that we are measuring the same sources from different angles, the same cannot be said for fMRI. As you will learn in a later section, the generators and time resolution of fMRI signals are very different to those of EEG and MEG, and there is still no clear model of how the electrophysiological sources of EEG and MEG relate to the metabolic fMRI signal. It cannot be generally assumed that activation 'hot spots' in fMRI are also visible in EEG/MEG, or vice versa. Even if there is some correspondence, brain areas shown as activated in fMRI maps may activate at different time points in EEG/MEG. However, some successful combinations of EEG/MEG with fMRI have been achieved for relatively simple perceptual processes. These are usually reflected in early EEG/MEG signals, and can be localized to a few focal brain areas.

In the example presented in Figure 14.8, the combination of EEG and MEG for the estimation of early visually-evoked responses is compared to fMRI data. If you have not read the section on fMRI yet, don't worry – for the moment, assume that fMRI localizes brain activity with greater spatial accuracy than EEG/MEG, and that for these specific examples it can serve as a kind of 'gold standard'. Simple visual stimuli ('Gabor patches') are shown in the rightmost column; these were flashed in the four quadrants of the visual field (upper left, upper right, lower left, lower right). It is known that such stimuli activate characteristic retinotopic areas of limited extent (~1 cm^2) in visual cortex in the occipital lobe (cortical area V1). Changing the stimulus location therefore changes the cortical response location in a predictable way. For example, visual stimuli presented in the left upper visual hemifield will activate V1 in the right lower cortical hemisphere. This known topography allows for response patterns to be compared across different localization techniques. The expected pattern of activation is clearly present in the fMRI data. This technique provides millimetre spatial resolution, but does not tell us whether this activation reflects early or late processes. EEG and MEG show a similar pattern around 74 ms after stimulus onset – this is about the time when visual information from the retina first reaches visual cortex. Localizations for EEG and MEG separately show approximately the same patterns, although EEG alone is quite far off in one case. The combination of EEG and MEG results in a pattern that closely matches that obtained using fMRI. This suggests that the fMRI pattern reflects an early processing stage in visual cortex. This is a simple sensory paradigm – the correspondence among imaging modalities may not be so clear for cognitively more complex tasks.

2.4 Statistical analysis of EEG/MEG data

The statistical analysis of EEG/MEG data is as yet not fully standardized, and the strategies researchers employ can vary considerably depending on their hypotheses. The most frequent approach is 'analysis of variance' or ANOVA. The signal amplitudes for different electrodes, sensors, or dipoles are considered as dependent variables, and recorded for each subject (just as we might measure mean reaction times or error rates per subject in behavioural experiments).

Imagine a so-called **oddball** experiment (the 'oddball paradigm' is discussed in Section 2.5), in which subjects are presented with two different tones: one tone of 600 Hz, occurring 90 per cent of the time, and another tone of 1000 Hz, occurring only 10 per cent of the time. Subjects are required to press different buttons in response to the frequent and rare tones. Does the maximum amplitude of the ERP in a certain time range change as a function of the probability of the tone? In this case, one would first have to decide which time range to analyse. In this kind of oddball experiment, it is known that rare tones produce a so-called **P300 component**, i.e. a positive deflection of the ERP around centro-parietal electrode sites, which usually reaches its maximum around 300 ms after stimulus onset. It would therefore be reasonable to look for the largest peak around 300 ms at the electrode called 'Pz', and to extract the amplitudes at this electrode and this point in time for all stimuli, for all subjects. These values would be entered into the ANOVA as the dependent measure. 'Tone rarity' would be a single, two-level factor allowing for a comparison between values obtained for the presentation of the rare tones and those obtained for the frequent tones. In general, the result is that the amplitude of the P300 is larger for rare tones.

Does only the amplitude change with the tone's rarity, or could there also be a change in the location of the peak evoked response? In other words, could the rare tones produce larger amplitudes in either the left or right hemisphere compared with the frequent ones? In this case, one would have to extract amplitude values for at least two electrodes, one in the left and one in the right hemisphere (e.g. P5 and P6, left and right of Pz). The ANOVA now has two factors: 'tone rarity' (rare and frequent) and 'laterality' (left and right). If the laterality changes with tone rarity, one should obtain an interaction between rarity and laterality. In the case of simple tones, such an interaction is generally not found, although it might be expected for more complex (e.g. language-related) stimuli. From our discussion on source estimation, however, it should be clear that the distributions of the electric potentials or magnetic fields around the head do not allow clear inferences about the sources in the brain. If hypotheses about brain areas (e.g. about lateralization) are important, then source estimation, or a different imaging method, should be employed.

2.5 The auditory oddball paradigm

The oddball paradigm, in which subjects are exposed to a stream of 'standard' stimuli occasionally interrupted by 'deviant' stimuli, is one of the most widely used paradigms in cognitive electrophysiology. This paradigm is easily tolerated even by severely impaired patients, and the corresponding ERP (to a lesser extent ERF) effects have been studied intensively during the last decades. Two major components are usually observed in ERP and ERF studies using oddball paradigms:

1. The **mismatch negativity** or **MMN**: an early component (between approximately 100 and 200 ms) that is observed even when the participant is attending elsewhere (e.g. they are watching a silent movie), and is generally interpreted as a 'pre-attentive change detection mechanism'.
2. The **P300**: a component peaking around 300 ms that is generally larger when target stimuli (e.g. deviants) have to be actively detected in a stream of standards.

The data presented in Figure 14.9 are taken from Opitz *et al.* (1999). The same 16 subjects participated in EEG as well as in a functional magnetic resonance imaging (fMRI) experiment. ERPs were recorded from 120 electrode sites. The stimulus set consisted of two 200 ms tones (600 Hz and 1000 Hz). The experiment was divided into 12 'standard' blocks, in which only the low tone was presented 24 times, and 12 'deviant' blocks, in which 16 low and 8 high tones were presented. This design was chosen in order to run the same experimental paradigm both with EEG and fMRI. In different runs, subjects were either told not to attend to the stimuli, but to watch a cartoon video instead ('unattend' condition), or to silently count the high tones ('attend' condition).

The detection of a change within a sequence of stimuli elicits an MMN shortly after 100 ms, independently of where attention is directed. Dipole modelling, where the position of the dipoles is based on the position of activations observed in the companion fMRI experiment, suggests that the sources of this MMN component are localized in left and right superior temporal gyri (STG). The later P300 is elicited when the deviant stimuli are relevant to the performance of a task, and the generators were estimated in left and right transverse temporal gyri (TTG). It should be noted, however, that the relationship between EEG/MEG sources and fMRI/PET activity is still not clearly understood. It is generally not straightforward to constrain EEG/MEG source localization by fMRI/PET results. The development of corresponding methods and their empirical evaluation is currently one of the most exciting areas in the field of neuroimaging.

SUMMARY OF SECTION 2

- EEG and MEG record the electrical signals of the brain millisecond by millisecond, and are therefore sensitive to changes in the temporal pattern of perceptual and cognitive processes.
- They do this only indirectly, which means that the active location in the brain can only be estimated.
- Data are usually analysed by computing 'event-related potentials/fields' (ERPs/ERFs): data segments are averaged across a large number of trials belonging to the same stimulus category. Differences between different averages may reveal the signals that are specific to a certain cognitive process.
- The same averaging procedure can be applied to frequency power spectra, which may reveal evoked and induced oscillatory activity.
- EEG and MEG signals are, in principle, generated by the same sources inside the brain, but the two imaging methods show different patterns of sensitivity. EEG 'sees more' than MEG, but MEG localizes better what it can see.
- Because of their very different time-scales, and the very different sources of the signal, a combination of EEG/MEG and fMRI neuroimaging modalities must be considered with great care, and may only be possible in very specific cases where the neuronal sources are already well known.

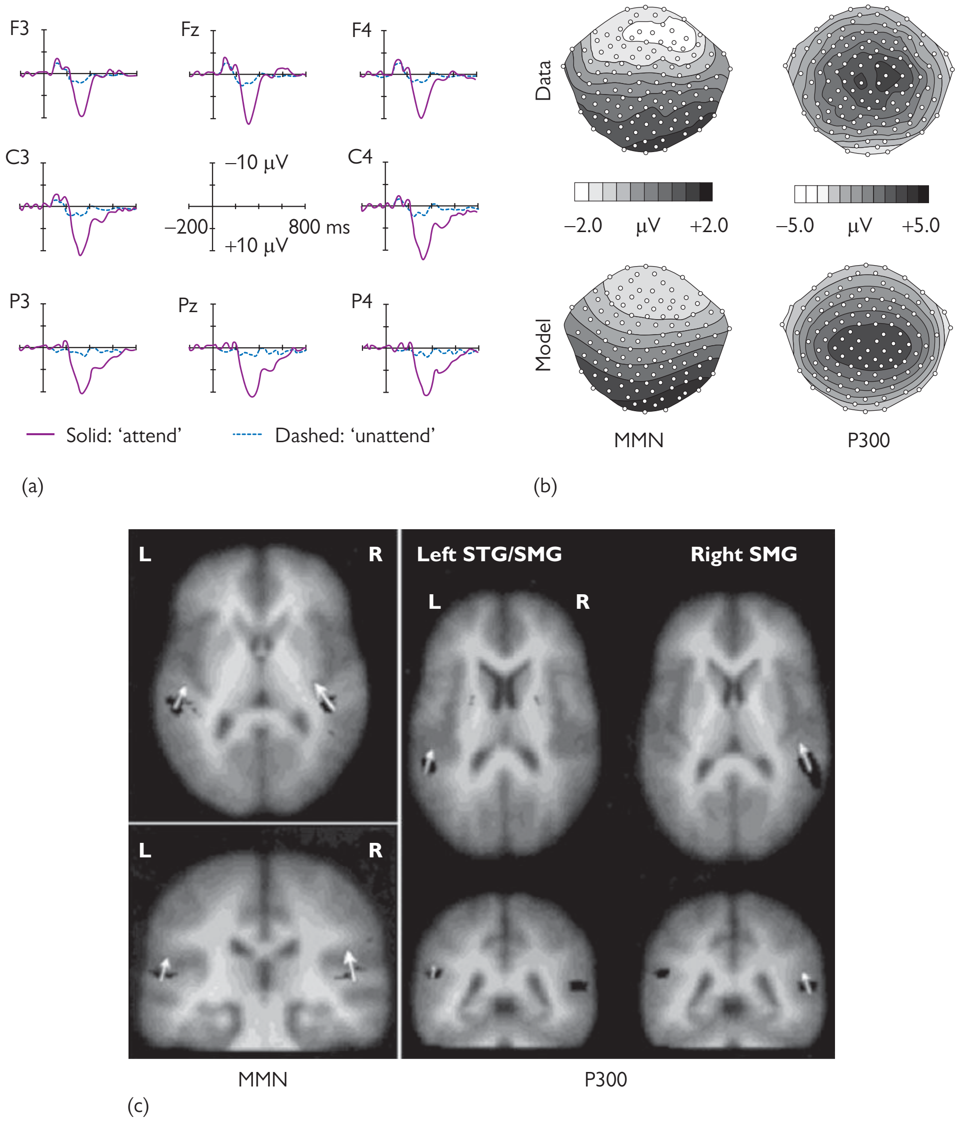


FIGURE 14.9 The auditory oddball paradigm: a continuous sequence of 'standard' stimuli (e.g. '… beep-beep-beep …') is occasionally interrupted by a 'deviant' stimulus ('… beep-boop-beep …'). This deviant stimulus evokes two distinct components in the ERP signal: the early 'mismatch negativity' (MMN) and the later P300. The P300 is usually bigger (or only present) when participants are required to actively attend to the stimuli (and for example respond to them, rather than just passively listen). The study presented here measured brain activity using ERPs and fMRI in the same group of 16 participants. (a) ERP difference waveforms ('standards' (low tones) minus 'deviants' (high tones)) averaged across subjects. The mismatch negativity (MMN) is reflected in a negative-going signal at frontal electrode sites (e.g. at electrode Fz around 150 ms) in both conditions, and the P300 as a large positive-going signal in the 'attend' condition only at posterior electrode sites (e.g. at Pz after 300 ms). (b) Topography of the MMN (left) and P300 components (right). Both the original data (top) and the prediction of the dipole models (bottom) are shown. (c) Two dipoles fitted to the ERP data explained up to 95 per cent of the data for both the MMN and P300. Dipole locations (as indicated by arrows) were consistent with fMRI activation clusters. STG: superior temporal gyrus; SMG: supramarginal gyrus. *Source: Based on Opitz et al., 1999, Figures 1–5*

3 TECHNIQUES BASED ON METABOLISM AND BLOOD SUPPLY

The brain requires a nearly constant supply of oxygen and glucose in order to function properly. Both glucose and oxygen are carried to the brain in the blood supply, via an array of arterial vessels. Although the brain accounts for only about 2 per cent of the body's mass, it consumes about 20 per cent of the body's glucose and oxygen, and receives about 20 per cent of its blood supply. Although the relationship is still not completely understood, activity in neurons is closely linked to glucose consumption, oxygen consumption, and blood flow. Blood in the brain goes to where it is needed: as a local group of neurons becomes active, blood flow in that region increases, as does the amount of oxygen in the blood in that region (**blood oxygenation**). Thus, imaging techniques that are sensitive to blood flow and blood oxygenation can be used to study 'the brain at work'. **Functional magnetic resonance imaging (fMRI)** is used by psychologists to measure how the concentration of oxygen in the blood changes in different parts of the brain as cognitive state changes. **Positron emission tomography (PET)** can also be used as a tool to measure how blood flow in different parts of the brain changes as a function of cognitive state. Before the advent of fMRI in 1992, PET was the main way to measure patterns of oxygen metabolism in the human brain. With the growth of fMRI, this type of PET study has become less popular due to the (albeit small) dose of ionizing radiation required, and its generally poorer temporal and spatial resolution compared to fMRI. It is still an important method for studying the metabolism of compounds other than oxygen in the brain (for example, compounds related to neurotransmitters). **Single-photon emission computed tomography** is much like PET, with even poorer spatial resolution. Finally, **near-infrared spectroscopy** uses long-wavelength light to study the oxygenation state of blood in the brain, and is commonly used with infants due to its ease of use and safety. We first discuss some of the most common designs used in neuroimaging experiments (particularly for fMRI, by far the most common metabolic technique in cognitive neuroscience), before explaining how the **dependent measure** in these designs (signal derived from blood flow, blood oxygenation, or other metabolic activity) is obtained in fMRI, PET, SPECT, and NIRS.

3.1 Experimental design

Neuroimaging experiments are designed to assess relative changes in brain activity (for example, indexed by regional changes in blood oxygenation in fMRI) consequent upon changes in stimulus characteristics or task demands. The two most commonly used approaches are subtractive designs and correlational (or parametric) designs.

3.1.1 Subtractive designs

In 1969, the Dutch physiologist Franciscus C. Donders proposed a general method to measure thought processes based on the assumption that tasks can be broken down into component processes (Donders, 1969). One comparison Donders made was between the reaction time to one stimulus presented by itself (simple reaction time) and the reaction time needed to identify a stimulus before responding. Donders believed that the subtraction of the simple reaction time from the identification reaction time would yield the time required for stimulus identification itself. Figure 14.10 gives a similar example in which the time taken to press a key in response (R) to a light coming on (the stimulus, S) (Task 1) can be subtracted from the time needed to press the key only when a light of a particular colour is illuminated (Task 2). Assuming that the identification component in Task 2 can be 'slotted in' between the two components in Task 1, with these other components remaining the same, then the subtraction reveals how long the process of identification takes.

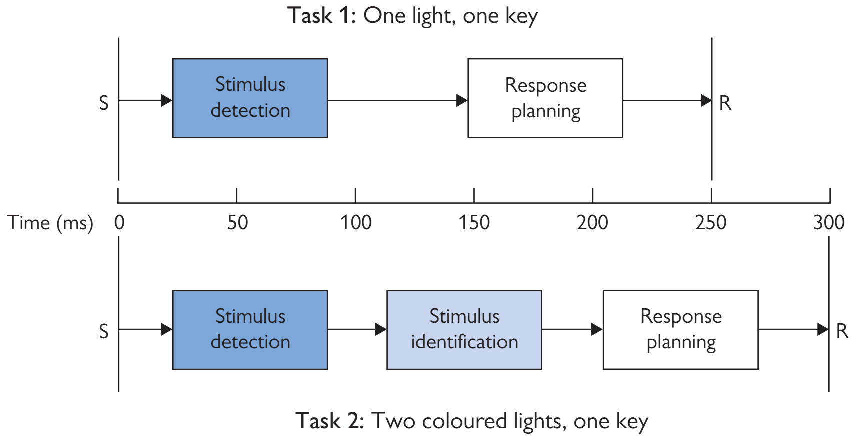


FIGURE 14.10 Cognitive subtraction. Task 1 involves responding to the detection of a light; Task 2 involves responding to the detection of a light of a particular colour.

The same subtractive logic is often used in the design of neuroimaging experiments. That is, if two conditions, *A* and *B*, are identical except that *A* requires a particular cognitive function f while *B* does not, then one can identify those cortical regions that are involved in f by subtracting the pattern of regional cerebral blood flow (rCBF) observed during *B* from that during *A*. This approach is essentially an extension of Donders's method (Raichle, 1998).

In general, the conditions that are being compared will differ in one of two ways. The stimuli presented during the two conditions may be identical but the participants may be required to perform different tasks, or the participants may be presented with different kinds of stimuli in each condition but the task that they are required to perform may be identical. Box 14.1 gives examples of these two approaches.

An elaboration of the idea of cognitive subtraction is the hierarchical subtractive design. In studies of this type, conditions are successively 'built up' from cognitive or perceptual components that are assumed to be additive. For example, in a study of speech perception (see Figure 14.11), rest (lying quietly) could be compared with listening to noise bursts, which would activate the auditory system but not the language system. Listening to words and listening to noise bursts would both activate the auditory system, but words would also activate parts of the brain involved in perceiving and understanding language. Repeating words would activate all these regions, and in addition regions involved in speech output. Generating synonyms to a heard word, as discussed in Box 14.1, would activate regions involved in semantic search and word retrieval, in addition to areas involved in hearing and repeating words.

Another elaboration of the subtractive approach is the factorial design, such as that used in behavioural research. We have already encountered a simple factorial design in the English/French PET study of Klein *et al.* (1995) (see Box 14.1). If we take the two conditions of repetition, and the two conditions of synonym generation, we have described two levels of two factors: Task (repetition or generation) and Language (English or French). As in a conventional behavioural experiment with a 2 × 2 factorial design, the data can be analysed to address questions about both main effects (what brain areas are active whenever French is used compared with English, across both tasks? What brain areas are more active during synonym generation than repetition, regardless of language?) and interactions (are there areas in which the activation difference between synonym generation and repetition is greater for French?). Factorial designs are powerful because the experimenter can look at two or more independent variables simultaneously, and also assess whether or not interdependencies exist between these variables.

3.1.2 Correlational/parametric designs

Correlational/parametric designs are used to examine the relationship between a stimulus parameter (such as word presentation rate) or a behavioural response (such as reaction time) and brain activity. Correlational designs

BOX 14.1 METHODS: Subtractive designs

Klein and colleagues (1995) were interested in whether different brain areas are recruited when volunteers use a language they are proficient in but acquired relatively late in life (in this case, French) compared with their native language (English). Participants heard words in English and French and were asked either to simply repeat what they had heard, or to generate synonyms of the words they had heard (in both languages). This latter task, in addition to recruiting areas involved with speech perception and production (as in the repetition tasks) would also recruit areas involved in semantic association and word retrieval and selection (as volunteers 'look for' and select a suitable synonym).

Comparing synonym generation to word repetition, within each language, is an example of the 'same stimuli, different task' approach. In both tasks, subjects heard single words, which they simply repeated in one condition, and to which they generated synonyms in the other. In both English and French, this comparison revealed activation in prefrontal cortex, indicating the importance of this region in the processing of word meaning and in response selection.

Comparing word repetition in the two languages is an example of the 'different stimuli, same task' approach. A small area in the left basal ganglia was observed to be significantly more active when volunteers repeated French compared with English words. This is interesting, since the basal ganglia are known to be involved in skilled motor action, including articulation of words. Perhaps even when English speakers are proficient in French, their motor apparatus has to work 'harder' to pronounce French words.

are elegant because they avoid many of the complexities of interpretation intrinsic to subtractive designs, including uncertainty about the assumption of additivity of cognitive processes (see Friston *et al.*, 1996). Correlation can also be used to assess brain connectivity and how it changes as a function of cognitive state. For example, signal change in one point in the brain can be used to predict activity elsewhere in the brain. If two areas show correlated activity, it is likely that these areas are connected in some way (directly or via another region). As cognitive state changes, so can connections among areas, reflecting recruitment of different functional networks.

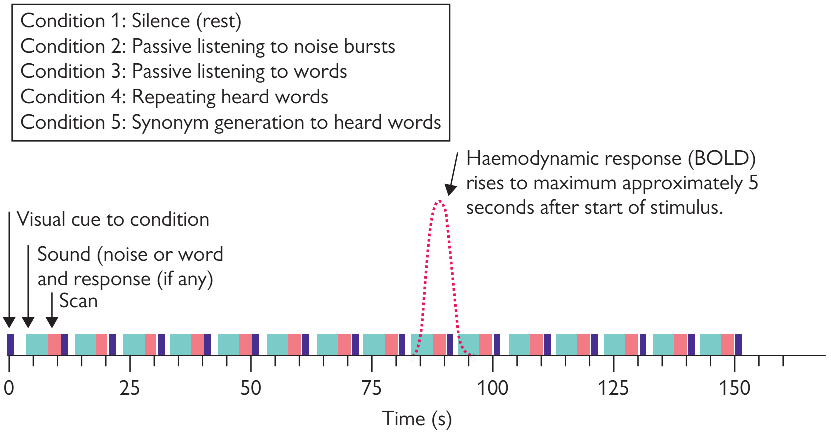

Condition 1: Silence (rest)
Condition 2: Passive listening to noise bursts
Condition 3: Passive listening to words
Condition 4: Repeating heard words
Condition 5: Synonym generation to heard words
Haemodynamic response (BOLD) rises to maximum approximately 5 seconds after start of stimulus.
Visual cue to condition
Sound (noise or word and response (if any)
Scan
0
25
50
75
100
125
150
Time (s)

FIGURE 14.11 Example of an fMRI experiment (a study of verb generation) employing a hierarchical subtractive design. In the 8-second gap before each 2-second whole-brain scan, volunteers see a cue on the screen (one of 'rest', 'listen', 'repeat', or 'generate') and then: hear nothing (in condition 1); hear a noise burst (in condition 2); or hear a word (in conditions 3–5). If the instruction is 'rest' or 'listen', volunteers lie quietly; if the instruction is 'repeat' they repeat the word that they heard; if the instruction is 'generate' they produce a synonym for the word that they heard. Each condition is performed 20 times in a pseudorandom order, so that each occurs ten times in the first half of the experiment and ten times in the second half.

We now turn our attention to the fMRI technique – what its advantages and disadvantages are, and how data are processed and analysed – before turning to the other, less common, imaging methods of PET, SPECT, NIRS, and finally TMS.

3.2 fMRI: functional magnetic resonance imaging

MRI is the general name for a broad class of techniques that use magnetic fields and radio waves to image the structure, organization, and function of biological systems without injecting radioactive tracers. Two of these techniques are used commonly in psychology and cognitive neuroscience. Structural MRI reveals the anatomical structure of the brain (see Figure 14.1), whereas functional MRI (fMRI) measures regional levels of blood oxygenation, which are assumed to reflect the volunteer's cognitive state. Since its discovery in 1992, fMRI has become a very popular tool for studying the 'brain at work'.

3.2.1 How MRI works

If you have ever played with magnets, you will know that magnetic fields are not stopped by materials that are electrically insulating. For example, a magnet under a thin wooden table will move a paperclip on its surface; similarly, magnetic fields pass easily through the skull. A large cylindrical magnet around a volunteer's head and upper body creates a strong magnetic field, with a direction that is parallel to the volunteer's body. Other magnets are applied so that the strength of this field changes systematically across the head. Various atomic nuclei, particularly the proton nucleus of the hydrogen atom (which is very common in the brain, since there are two atoms of hydrogen in every water molecule and the brain is mostly water), align themselves with the magnetic field. These hydrogen protons, which spin anyway, precess (like the wobble of spinning tops) about the axis of the applied magnetic field at a characteristic frequency that depends on the strength of the field that they are in (see Figure 14.12). Initially, the protons precess (wobble) out of synchrony with each other. When radiofrequency energy is applied, the protons suddenly become synchronized in their movements, but because they all precess at different rates, they drift out of synchrony again quite soon. However, as this happens they release energy. The frequency of the energy they release depends on the field strength they experience; that is, their position in the head. Sensors just outside the head detect the frequency of these energies, and a computer can be used to reconstruct an image of the distribution of protons, and with it, different tissues in the brain (e.g. grey matter, white matter, blood, cerebrospinal fluid). This tissue contrast is coded in greyscale in the images.

Functional MRI (fMRI) relies on the different magnetic properties of oxygenated and deoxygenated blood, and allows scientists to observe how the relative concentrations of oxygenated and deoxygenated blood change with cognitive state. Oxygen is carried to the brain in haemoglobin molecules in red blood cells. The bright red blood in the arterial supply relinquishes its oxygen to brain tissue in capillary beds (where oxyhaemoglobin is converted into deoxyhaemoglobin). Purplish deoxygenated blood flows away from capillary beds in the draining venules and veins. As neural activity increases in a part of the brain, more oxygen is extracted from the blood, but blood flow and blood volume in that part also increase. The blood volume increases more than is necessary – it overcompensates for the increase in oxygen consumption.

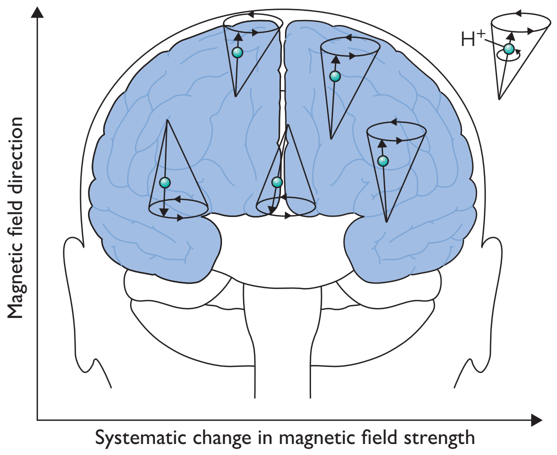


FIGURE 14.12 Hydrogen nuclei (protons) aligning in a strong magnetic field in the head of a volunteer. Each proton spins about an axis; in turn, the axis also wobbles or precesses around the axis of the magnetic field.

(Imagine every fire engine in London arriving on the doorstep of a chip-pan fire in Notting Hill!) More oxygenated blood is supplied than is needed, which has the effect of decreasing the concentration of deoxyhaemoglobin in the system of venules and veins draining the patch of active tissue (see Figure 14.13(a)). Since deoxyhaemoglobin and oxyhaemoglobin have different magnetic properties, this change in concentration alters the magnetic field strength in the region, which is reflected in the patterns of energy released by protons during scanning. This effect is known as **blood oxygen-level dependent (BOLD) contrast**. Thus the fMRI signal is only a secondary consequence of neural activity and does not measure neural activity directly. The **haemodynamic response**, which is the change in BOLD that occurs with a change in neural activity in a particular brain region, happens much more slowly than activity in neurons does. It takes 4–5 seconds to develop after neural activity increases, and takes as long again to decay after neural activity stops (see

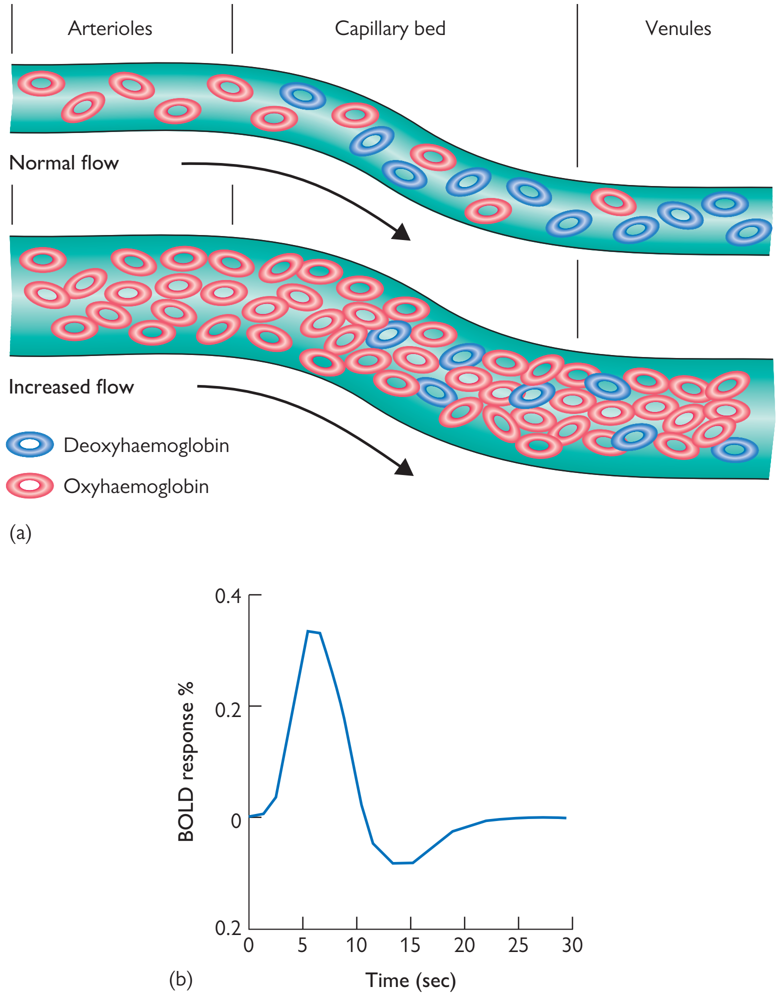


FIGURE 14.13 Blood oxygen-level dependent (BOLD) contrast and the haemodynamic response function. (a) Functional MRI scans are exquisitely sensitive to subtle fluctuations in magnetic field strength. Oxyhaemoglobin, found in arterial blood, and deoxyhaemoglobin, found in venous blood, have different magnetic properties. When their relative proportions in the blood change, as happens when a patch of neural tissue becomes active, the resulting change in the local magnetic field can be detected. (b) The metabolic change described in (a) is not instantaneous: it takes several seconds to develop (haemodynamic lag) and equally long to settle back to baseline. The example here sketches, for a point in auditory cortex, what the percentage change in BOLD (from a resting level) might be as a function of time since hearing someone snap their fingers (at time 0). This is called the haemodynamic response function (HRF). *Source: (a) Adapted from P. Jezzard, http://www.fmrib.ox.ac.uk/fmri_intro/brief.html*

Figure 14.13(b)). Thus, when the experimental condition changes, BOLD response takes 4–5 seconds to change also. This is called the haemodynamic lag. This sluggish response limits our ability to know exactly when a bit of brain tissue actually became active after stimulus presentation, but since the lag is relatively constant, the timing of two or more events relative to each other can be measured with greater precision; potentially within a few hundreds of milliseconds (Menon and Kim, 1999).

The spatial resolution of fMRI is mostly limited by the strength of the magnet used to generate the main magnetic field: the stronger the magnet, the finer the potential resolution. This is for two closely related reasons. First, stronger magnets are just more sensitive, so small signal changes (small areas of activated tissue) can be detected more easily, and two such small regions of activation can be more easily distinguished. Second, more sensitivity means that instead of picking up signals in veins (which produce large signal changes), signals in small venules and in the capillary bed can be detected. Since venules and capillaries are much closer to the active neural tissue than veins are – veins can be several millimetres away from actually active tissue – spatial resolution is improved. The fact that signals in veins can be mistaken for signals in brain tissue, unless precautions are taken, is one limitation of fMRI. Further limitations and advantages are detailed in the next sections.

3.2.2 Limitations

fMRI is sensitive to changes in the local blood supply when a population of neurons becomes active. It is only an indirect measure of neural activity, and since changes in blood supply happen slowly, temporal resolution is limited, and is not as good as with EEG and MEG.

We have already explained how, since large BOLD changes can occur in veins, apparent activation can be located at some distance 'downstream' from brain tissue that is actually active. Without knowing where veins are located in each subject (like brain anatomy, the organization of blood vessels, particularly on the venous side, can be very different from person to person), venous activation can be mistaken for tissue activation. This obviously makes it difficult to infer confidently that the tissue in a region of significant signal change is actually involved in the cognitive process being studied (the real activated tissue may be up to a centimetre or more 'upstream'). There are several ways of getting around this problem. First, more powerful magnets, or different imaging protocols, can be used to improve signal strength in venules and capillaries. Second, MRI can also be used to make maps of blood vessel distribution (angiograms) in each individual: these maps can be compared to patterns of activation to identify and eliminate activation in vessels.

Several other limitations of the fMRI technique are based on the physics of the BOLD signal. BOLD is a special case of a subtle fluctuation in local magnetic field strength, which can be detected with specialized imaging protocols. Such imaging protocols are necessarily sensitive to all subtle changes in local field strength, including those caused by the brain's structure itself, not just by the state of the local blood supply. For example, air-filled pockets in the head (like the sinuses) have a very different effect on local magnetic field strength than do brain tissues; thus the magnetic field changes quite dramatically at the front of the head where the sinuses are. Because fMRI actually requires a quite uniform magnetic field, these relatively dramatic changes in field strength produce decreased signal (and sometimes no signal at all) in affected regions. For this reason, inferior frontal and temporal cortex, close to such air pockets, can be hard to image. Unfortunately these regions are interesting to psychologists since they are implicated in decision-making behaviour, the sense of smell (orbitofrontal cortex), emotion (amygdala), memory (hippocampus), and language (lateral temporal cortex). An inability to 'see' these areas properly using fMRI makes it difficult (but not impossible; again, strategies to minimize this problem are in quite common use) to use as a research tool in these areas.

Fluctuations in field strength (field inhomogeneities) can also produce geometric distortion in images, apparently 'stretching' some parts and 'squashing' others, resulting in an image that is quite a different shape from the real brain. A popular way to 'correct' images for this problem is to explicitly measure the magnetic field inhomogeneities in each volunteer, and then use this information to compensate for the geometric distortions.

Other limitations of fMRI are purely practical. First, the strong magnetic field means that volunteers with ferromagnetic implants (such as pacemakers, pins to repair broken bones, or most types of cochlear implant) cannot be scanned safely. It also makes for a difficult experimental situation, since totally magnet-compatible stimulus-delivery systems (video projectors, headphones, response boxes) must be used. However, most fMRI systems are now equipped with a video projector that can present a picture (usually the screen of the computer controlling the stimulus delivery), headphones through which sounds can be presented, and response devices such as a button box or joystick.

When the scanner is acquiring data it is very noisy – noise levels up to 110 dB are common. This is loud enough to damage unprotected ears. This noise is unpleasant, and even when partially blocked with earplugs it can be distracting. Given this high level of background noise, special care must be taken when designing studies using sound. The most common approach is to scan intermittently, not continuously, alternating several seconds of sound presentation with 1 or 2 seconds of scanning (see Figure 14.11). In this way, sounds are presented in a quiet background, and the delay in the haemodynamic response (see Figure 14.13(b)) means that a scan taken 4–6 seconds after stimulus presentation will optimally detect elevated signal due to that stimulus.

The data quality of fMRI is seriously compromised by subject movement, so subjects must stay perfectly still during scanning. This also means that they cannot speak, so a task requiring speech or other movements of the head is not practical. Most fMRI tasks involve either key press or joystick responses. The restriction on movement means that patient groups with movement disorders can be difficult to scan. Finally, some subjects cannot tolerate the confined space of the MRI scanner.

3.2.3 Advantages

fMRI can be repeated within subjects, so it can be used to study processes that change over time, for example as a function of development, or as an individual recovers from injury or stroke, or as they acquire a new skill. It has quite good temporal and spatial resolution, and these both continue to improve as techniques are refined.

An fMRI scan can produce an image of BOLD signal, covering most or all of the brain, as fast as every second. If an experiment is targeted at a particular brain area (the auditory cortex, for example), then a small brain volume consisting mostly of the area of interest can be scanned, and **acquisition time** (the time required to take one scan) can be reduced even further. In principle, fMRI can produce an image that distinguishes structures less than a millimetre apart, and this potential is being realized. For example, in visual cortex, cells that respond preferentially to lines of a given orientation are arranged into columns, which in the cat are half a millimetre wide and approximately 1 mm apart. Kim and colleagues (2000) were able to use fMRI to map such orientation columns in the visual cortex of the cat.

In summary, fMRI is an increasingly common tool that psychologists use to determine how different parts of the brain are specialized for different cognitive processes. Although there are several limitations to the technique, it is a field that is growing and developing rapidly. We shall now discuss how imaging data must be processed before they can be analysed (the treatment of functional PET and SPECT data is quite similar).

3.2.4 Processing and analysis of fMRI images

The analysis of imaging data is a multi-stage process (see Figure 14.14). First, the scans must be **reconstructed**. The raw data collected in PET, SPECT, and fMRI do not resemble an image. A mathematical transform, such as Fourier analysis, is required to turn the data into a set of images: **the image time series** (the sequence of images acquired over time). Additional steps are then often required for fMRI data to correct for artefact and distortion. We now review the most important processing steps required for reconstructed images to be prepared for the stage of processing that is closest to the researcher's heart – the statistical analysis that reveals where significant activation related to experimental variables occurred.

First, although movement of the head is generally restrained in some way (by lying in a moulded helmet, or with a headpiece), some movement is nevertheless

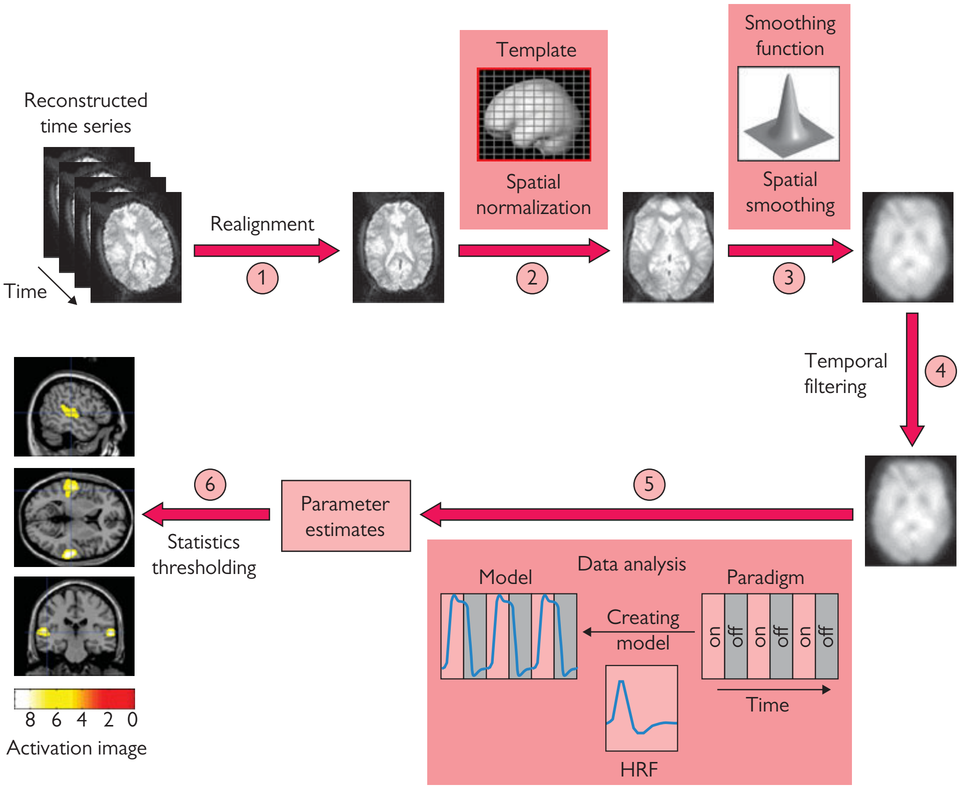

Reconstructed time series
Time
Realignment
1
Template
Spatial normalization
2
Smoothing function
Spatial smoothing
3
Temporal filtering
4
5
Parameter estimates
6
Statistics thresholding
Data analysis
Model
Paradigm
Creating model
on
off
on
off
on
off
Time
HRF
8 6 4 2 0
Activation image

FIGURE 14.14 Typical steps of image processing and analysis of fMRI data. Processing of PET and SPECT data is very similar up to the analysis stage. Because images are not acquired as quickly with PET and SPECT as with fMRI, the analysis of such data does not need to take the haemodynamic response function (see Section 3.2.1) into account.

inevitable over the course of a scanning session (several minutes to a few hours). Any movement means that the position of the brain within the images will vary over time. Thus, the images need to be realigned, bringing them all back into registration and correcting for head movement.

One concept that is almost universally used in functional imaging is that of a standard space – a standard space provides a way to locate positions in the head (in three dimensions) and allows comparison among different brains (see Figure 14.15). First, an imaginary line is drawn between the anterior commissure (AC) and the posterior commissure (PC) – two small brain structures that are in very similar positions from person to person. This is the y-axis, and it defines front–back (see Figure 14.15(a) and (b)). Another axis, x, is perpendicular to y, also goes through the AC, and defines left–right (Figure 14.15(b)). A final axis, z, is perpendicular to the first two, also intersects the AC, and defines up–down (Figure 14.15(a) and (b)). All three axes intersect, or pass through the origin, at the AC. These axes define the standardized stereotaxic coordinate system, sometimes called the Talairach coordinate system after an influential brain atlas (Talairach and Tournoux, 1988). Any location in the brain can be indexed by values of these three coordinates, using units that are approximately a millimetre (Figure 14.15(c)).

Spatial normalization (see Figure 14.14) refers to the process by which brain volumes are changed in shape, size, and orientation so they approximately match that of a template, in standard space. A template is usually an average of many representative individual brain volumes. This is done so that brains of different subjects (who will have brains of different shapes and sizes) can be easily compared, and so data can be averaged across subjects. It also allows a researcher to match up low-resolution functional data with a high-resolution anatomical image, so activations can be precisely located with respect to brain

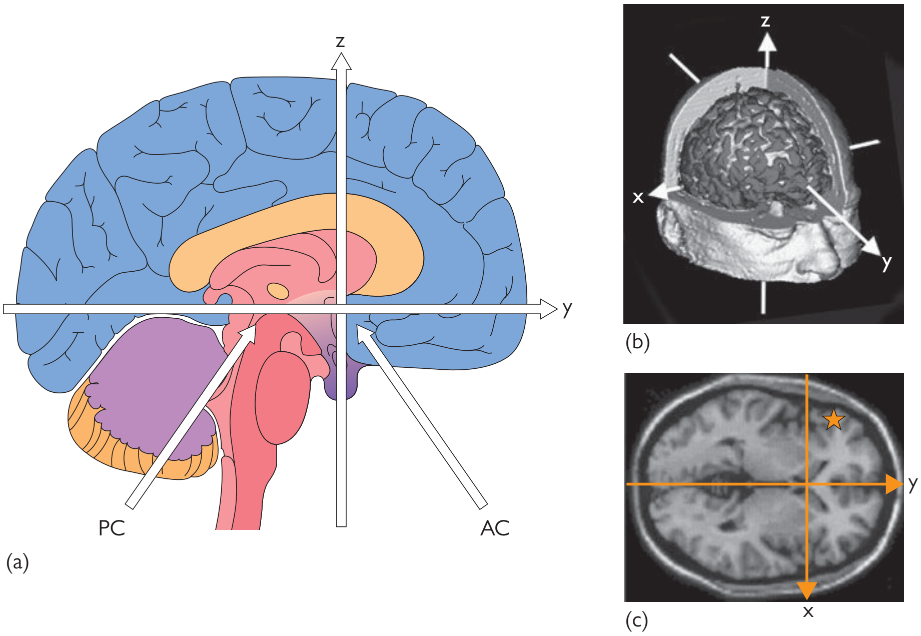


FIGURE 14.15 Standardized stereotaxic coordinate system, with perpendicular x-, y-, and z-axes (see (a) and (b)). The star in (c) is in the location (x, y, z) −45, +15, +10. This is approximately 4.5 cm left of the mid-line, 1.5 cm in front of the anterior commissure, and 1 cm above the AC-PC line. This point always falls in the left inferior frontal gyrus in normalized brains.

cortical anatomy. Anatomical landmarks, such as the anterior commissure, the Sylvian fissure, and the interhemispheric fissure, will be in an (almost) identical position in all normalized brains. (Brains are all very different, however, and even after spatial normalization some anatomical variability remains.)

When brains are spatially normalized, a particular set of stereotaxic coordinates refers to the same brain region in everyone (see Figure 14.15(c)). This permits the researcher to use reference atlases, which are in the same space as the template and which give, for each set of x, y, z coordinates, detailed information about the brain at that point. For instance, atlases can provide information about the likely tissue type, anatomical structure, and what other brain regions might be connected.

It should be noted that sometimes researchers are interested in knowing more about what is different in the anatomy and pattern of activation across individuals, and are not so interested in what is common. In such cases, spatial normalization would remove some of the differences that interest the researcher, and so probably would not be used.

After normalization, data are usually spatially smoothed, meaning that the activation at each point is replaced by a weighted average of the activation measured not just at that point, but at all neighbouring points. This has three benefits. First, it improves power, making activations easier to detect. Second, it ensures that the data conform to the assumptions of the statistical model that will be used to determine the significance of the activations. Third, it compensates for the fact that the same activation might be in slightly different places across individuals, since structural and functional anatomy will not be identical even after spatial normalization.

The disadvantage of spatial smoothing is that it reduces spatial resolution, since the signal is ‘smeared out’ over adjacent regions. This effect is generally greater in PET data than in fMRI data: in PET data, two activations up to 2 cm apart can be smeared together to appear as one, whereas in fMRI, once activations are more than 1 cm apart, they usually appear as separate. The experimenter can change the amount of smoothing in order to examine whether what is apparently a single blob after smoothing might actually arise from two or more discrete but neighbouring locations.

The final step before data analysis is temporal filtering. As with EEG and MEG data (see Section 2.2.3), noise in fMRI data can be removed by filtering. For example, it is common to use a high-pass filter in order to remove slowly varying components in the time series, which may be due to physiological effects or to scanner-related drifts in baseline signal. Once this is done, the data are (finally!) ready for statistical analysis.

3.2.5 Data analysis

Predicting how signal will change over time, and testing such models

The most common way to analyse imaging data is to set up a model that predicts how the signal in the brain will change over scans if it is related to stimulus presentation or task performance. For the example in Figure 14.11 regions of the brain that are uniquely involved in synonym generation would be expected to show greater signal in scans 2 and 10 (when subjects were performing the synonym generation task) than in scans 3 and 6 (when they were repeating words, a condition that is closely matched in its other cognitive demands to synonym generation). Each point in the brain is tested separately; that is, at each point in the brain, paired t-tests (over subjects) would indicate whether the signal is higher during synonym generation than during word repetition.

In fMRI, since scans are usually taken every 1–4 seconds, the haemodynamic response function (HRF) (see Section 3.2.1 and Figure 14.13(b)) must be taken into account when predicting how measured activation will change over time, as a function of stimulus or task. The data analysis stage in Figure 14.14 shows a block design in which sound alternates between being 'on' for several seconds and 'off' for an equal period of time. Because of the HRF, the BOLD signal that would be predicted from this design would not increase until a few seconds into the 'on' periods, and then would stay elevated for a few seconds after the sound has been turned off. At each point in the brain, regression equations are used to test how well the model fits the observed signal change over scans. The goodness of fit is indexed by parameter estimates. The larger the parameter estimate at a given point in the brain, the better the model fits at that point.

Statistical tests are carried out at each of several thousand points in the brain, independently (200 000 points would not be unusual). If we set the significance threshold to 0.05, which is conventional for single statistical tests, then it will be practically guaranteed that one of the tests will reach significance by chance alone (i.e. it will generate a false positive or Type I error). Indeed, with so many tests, there are likely to be many Type I errors. How can we distinguish between these false positives and those comparisons where the signal was truly elevated for experimental reasons? We must correct for the fact that so many tests are being carried out: we need a multiple-comparisons correction. This can be done in several ways, but such statistical thresholding always has the effect of increasing the statistical threshold at which something is judged to be significant. The activation image (see Figure 14.14) shows which points in the brain are activated at a particular (specified) level of significance, usually $p < 0.05$ corrected for multiple comparisons.

Sometimes researchers have predictions about where in the brain they are expecting to find activation. In such cases, they do not need to correct for tests conducted over the whole brain, just for the tests conducted in the region in which they are interested. The statistical criterion for significance is therefore lower than if a correction had been applied based on tests over the whole brain. The smaller the region they are interested in, based on their hypotheses and on previous work, the smaller the correction needs to be. It is usually helpful in imaging studies to have anatomical hypotheses, based on existing literature, about where activation is likely to be observed when perceiving a particular kind of stimulus or performing a particular task.

Functional connectivity analysis

The traditional approach to the analysis of fMRI (and earlier PET) studies developed naturally from neuropsychology (see Chapter 13). Neuropsychologists are able, by studying patients with brain damage restricted to one part of the brain, to infer the functional specialization of that damaged area. Similarly, traditional approaches to imaging data analysis were focused on evaluating the functional specialization of discrete brain regions. However, since imaging data sets contain information about the moment-to-moment signal levels across the whole brain, they can also be analysed to reveal how signal in different areas is coupled; for example, how signal levels across multiple areas rise and fall in concert, from moment to moment (i.e. activity is correlated across regions, even within a particular condition). This is called 'functional connectivity' and analyses of such correlations are more and more popular since regions that show coupled activity probably belong to the same functional network. One of the most popular ways to measure such connectivity is in the resting state – to see how function is coupled across the brain when people are being scanned just at rest. This reveals an extensive interconnected set of brain regions often called the 'default network', which is thought to be important for internally generated cognition – for reflecting on the past, for planning the future, for managing social interactions, and for engaging in mental simulations relevant to one's life. Alterations in the default network have been noted with ageing and in a number of neurological disorders.

Pattern classification and 'representational similarity'

Another way to analyse imaging data is to interpret the spatial pattern of activity associated with a particular stimulus or condition (across the whole brain or across a specific region) as a representation of that stimulus or condition. Then researchers can ask whether the patterns associated with other stimuli or conditions are similar to that pattern, or different. In other words, whether there is representational similarity (see Figure 14.16). For example, Kriegeskorte

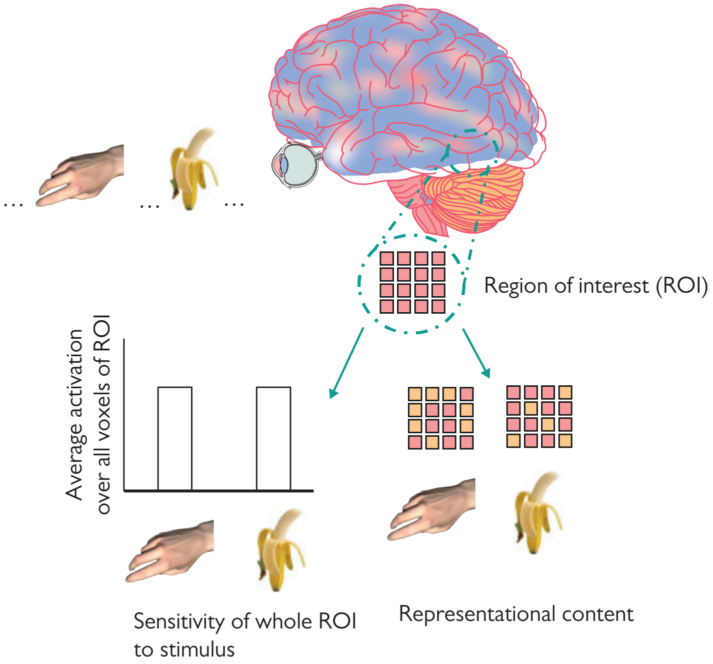

…
…
…
Region of interest (ROI)
Average activation over all voxels of ROI
Sensitivity of whole ROI to stimulus
Representational content

FIGURE 14.16 Whereas typical fMRI data analysis will tell you if a region is sensitive to a given type of stimulus, pattern-information analysis indicates how the region is involved, in a voxel-specific way. A specific ROI, here in inferotemporal cortex, can show similar sensitivity to two different classes of stimuli (here, body parts and fruit) but this same activation level results from two different patterns across the 16 voxels of the ROI. Pattern information analysis, or representational similarity analysis, assesses these multi-voxel spatial patterns, not the average activation revealed in a typical analysis. *Source: Based on the results of Kriegeskorte et al., 2008, and a diagram in Mur et al., 2009)*

and colleagues (2008) performed a study in which they examined whether monkeys and humans categorize objects in a similar way. They used electrophysiology in monkeys, and fMRI with representational similarity analysis, to show that in fact, inferotemporal cortex (IT) of the monkey and the human both exhibit similar response patterns. The IT cortex exhibits different patterns for animate and inanimate objects, and within these, patterns differ for different sub-categories (e.g. faces vs body parts; manmade objects vs fruit). Like connectivity analysis, this approach exploits the fact that imaging data sets contain information about multiple brain regions at every time point. These classification methods are based on the observation that, if the experimental conditions (which give rise to the mental representations of interest) can be classified at better-than-chance levels solely on the basis of the distribution of activity in the brain, then this activity pattern must carry appreciable information about the experimental conditions. This approach has been referred to as 'decoding' or even 'brain reading' (Cox and Savoy, 2003), and holds great promise for studying the nature of mental representations via fMRI.

3.3 PET: positron emission tomography

3.3.1 How it works

PET cameras are used to measure local concentrations of biologically important compounds (such as glucose, water, neurotransmitters, or drug compounds). Positron-emitting isotopes of oxygen, fluorine, nitrogen, carbon, and other elements are incorporated into glucose, water, or other neurochemicals (compounds used by the brain) using chemical reactions, and then these are introduced into the body. (Positrons are positively-charged electrons that are emitted from the nucleus of unstable (radioactive) isotopes, such as oxygen-15.) The emitted positron combines with an ordinary electron of a nearby atom in an annihilation reaction, turning the mass of the two particles into gamma rays (high-energy light particles) that are emitted at 180 degrees from each other. These photons easily escape from the human body, and can be recorded by a ring of radiation detectors: the PET scanner (see Figure 14.17). By recording the detection of these photons it is possible to monitor the distribution of neurochemicals, without disturbing their physiological behaviour, with a temporal resolution ranging from several seconds to several minutes, and with a spatial resolution of several millimetres.

3.3.2 Different kinds of PET study

PET is often used to study resting metabolism, that is to say, brain activity in the absence of any particular cognitive task. After being injected with the tracer compound, patients are scanned while lying quietly in

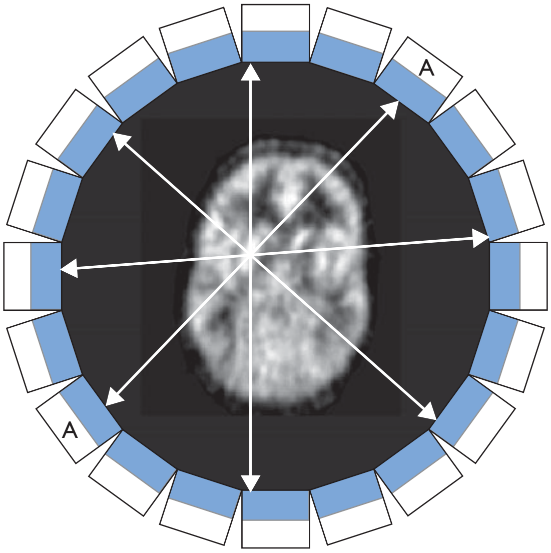


FIGURE 14.17 Coincidence detection in a PET scanner. When high-energy photons (gamma rays) are detected at opposite detectors (A), positron annihilation must have occurred somewhere along the coincidence line (the straight line connecting the two detectors). The intersection of many successively detected coincidence lines, observed over the course of a scan, pinpoints areas of relative concentration of the radioactive label. A computer uses the data gathered by the detectors to construct two- or three-dimensional images that show where the labelled compound (e.g. water (blood), glucose, or other neurochemical) accumulated in the brain.

the scanner. Such metabolic scans are useful for diagnosing brain disease: in some diseases, affected brain tissue is often less metabolically active than healthy tissue, appearing as 'dark areas' on the resulting images.

When psychologists use PET, however, they are generally interested in knowing where activity increases in the brain as a function of a particular cognitive task. These are called **activation PET** studies. Most activation PET studies use oxygen-15-labelled water as a tracer, since: (a) this isotope decays quickly and so disappears rapidly from the body; (b) blood is mostly composed of water; and (c) water crosses the blood-brain barrier easily. Labelled water is injected into the bloodstream of the volunteer or patient, usually in the arm, and diffuses rapidly throughout the blood supply as the volunteer performs a particular task (listening to a series of words, for example). As brain areas become active (such as auditory cortex, in response to spoken words), blood flow to those areas increases, and this is reflected in an increased concentration of the isotope in that region. Many gamma rays emanate from the region of concentrated isotope and are detected by the PET camera. Thus, activation PET, using labelled water, measures regional cerebral blood flow (rCBF). In most activation PET studies, 8–12 scans are taken, each approximately a minute long, with a 7–10 minute interval between scans; this is long enough for the tracer to disappear from the body before the next injection is given.

3.3.3 Limitations of PET

The most obvious limitation of activation PET is shared with fMRI (see Section 3.2.2): blood flow is only an indirect measure of neuronal activity, which is what researchers are really interested in. And changes in blood supply happen several seconds *after* the population of neurons under study has become active.

Furthermore, compared with fMRI, activation PET has many disadvantages that explain its lack of popularity in the last decade. Most countries, including the UK, have strict rules regarding the administration of ionizing radiation to ensure the safety of the general public. In the UK, the annual dose to members of the public is limited to 5 milliSieverts (mSv), comparable to a year of exposure to natural background radiation. Thus, in a typical activation PET study, a participant may experience 8–12 scans, and the total exposure across all of these must not exceed 5 mSv per year. More scans than this are not feasible, since the camera's sensitivity to blood flow depends on the size of dose given in any particular scan. If more scans were taken, then dose per scan would be even smaller. If too little isotope is given, then the camera does not detect gamma rays related to local accumulation of labelled blood.

Regulations in most countries also stipulate that PET cannot be used as a research tool with children or, in the UK, with women of childbearing age, and radiation dose limits make it difficult to repeat tests with individuals. This limits PET as a tool for investigating processes that occur over a period of time, such as patterns of brain activity before and after learning a skill.

PET is also expensive. In order to create the isotopes that are used as tracers, and to incorporate these into biologically important compounds, an array of sophisticated equipment is required. This equipment is expensive to both operate and maintain. Because most of the isotopes that are used in PET decay quickly, they have to be made in a **cyclotron** adjacent to the camera, and as soon as needed.

3.3.4 Advantages of PET

PET is a quiet imaging technique, and high-fidelity sound delivery can be easily used with it, making it a valuable tool in auditory research. Unlike fMRI, PET is compatible with the metallic and magnetic components contained in many implants, such as pacemakers and cochlear implants. Furthermore, unlike most fMRI systems, PET permits the researcher to be in close contact with the patient or volunteer: the volunteer is scanned for at most two minutes at a time, preceded and followed by contact with the researcher or radiographer. Also, PET scanners only encircle the head, leaving the body mostly open to the room, and are consequently less likely than MR scanners to provoke claustrophobic reactions.

Most importantly, PET is not restricted to oxygen as a tracer: other radioisotopes can be incorporated into other compounds that are active in the brain (neurochemicals) and that bind specifically to brain receptors of various types. For example, brain receptors for the neurotransmitter dopamine (**neurotransmitters** are chemicals that act as messengers between cells in the brain) are known to play a crucial role in many cognitive processes, in motor function, and in emotion. A compound called **raclopride**, incorporating a radioactive isotope of carbon, binds to a subtype of dopamine receptors. Another compound, **fluorodopa** (labelled with a radioactive isotope of fluorine) mimics a chemical precursor of dopamine. These compounds can be used to study the dopamine system both in normal individuals and in those suffering from disorders that affect it (Parkinson's disease, schizophrenia, depression, addiction). Many other compounds, specialized for study of other neurotransmitter systems, currently exist and more are being developed. This sort of research is not very common because it is very hard to do: complex chemical reactions are required to produce radioactive compounds that are specific to particular brain receptors.

Although tracers other than oxygen are usually used in resting metabolic PET studies, some recent experiments have used such tracers in activation studies. These experiments explore the effects of cognitive state on activity within neurotransmitter systems. For example, in one PET study, raclopride was given to eight men and they were scanned twice, once while playing a video game with a monetary reward, and once while staring at a blank screen (the control task). The results showed that playing the game was more associated with dopamine release and binding with dopamine receptors in a part of the brain (the **ventral striatum**) that is known from animal research to be involved in the appreciation of reward. Furthermore, this increase in dopamine release and binding correlated with level of performance achieved on the video game (Koepp *et al.*, 1998). Thus PET is a unique tool with considerable potential for mapping and studying neurotransmitter systems and neurochemically specific processes in both healthy volunteers and in patient groups.

3.4 SPECT: single photon emission computerized tomography

In SPECT, as in PET, a small amount of a radioactive compound is injected into the body. The compounds that are used generally mimic the action of one of the brain's neurochemicals, and so accumulate in the brain in areas where the metabolic process involving the compound of interest typically occurs. As with PET, gamma rays are emitted from the compound, and these are detected by a **gamma camera**. A computer is used to reconstruct the brain image in three dimensions. Like PET, this imaging technique is used to measure metabolism throughout the brain. Generally, SPECT tracers are more limited than PET tracers in the kinds of brain activity they can monitor. SPECT tracers also deteriorate more slowly than many PET tracers, which means that SPECT scans usually last longer than PET scans do, and much more time must elapse between scans to ensure that radiation levels fall to baseline. However, because SPECT tracers are longer lasting, compounds can be synthesized in centralized laboratories and then shipped to facilities where SPECT is used, so SPECT requires much less specialized technical equipment and staff than PET. While PET is more versatile than SPECT and produces more detailed images with a higher degree of resolution, particularly of deeper brain structures, SPECT is much less expensive.

SPECT has several clinical uses, including the determination of the brain area giving rise to epileptic seizures, and differential diagnosis of dementias (see McArthur *et al.*, 2011).

3.5 NIRS: near-infrared spectroscopy

If you have ever looked at your hand backlit by a strong light source such as a torch, you know that light passes through body tissue. NIRS uses near-infrared light (long wavelength, such as is used by warming lamps) to image the surface of the brain. This is a non-invasive technique because the light source passes through the scalp and skull, is reflected off the cortex, and the reflected light is picked up by a detector near the light source on the head's surface. Because blood flow and blood oxygenation alter the reflectivity of the cortex, brain activity is 'reflected in' (excuse the pun!) a change in the detected reflected light. NIRS does not, like fMRI and PET, have the ability to examine all regions of the brain at the same time, but can only be used to examine the function of a restricted region of cortex, and only near the surface of the brain, since the light does not penetrate far beyond the surface. This technique is non-invasive and safe, highly portable, requires (in comparison to other imaging techniques) relatively little equipment (see Figure 14.2(f)) and is relatively inexpensive. Furthermore, due to the relatively thin skulls of infants and children (compared to adults), which permit more light to pass, this technique is particularly well suited for use with younger participants.

SUMMARY OF SECTION 3

- fMRI is sensitive to changes in regional blood flow and blood oxygenation, indirect measures of neural activity that develop over 4–6 seconds.
- fMRI can be used to study regional brain activity but also, increasingly, how regions of the brain work together to represent stimuli or perform tasks.
- PET requires the use of ionizing radiation, but is an appropriate technique for anyone with metallic or magnetic implants (such as pacemakers or cochlear implants) who cannot undergo fMRI.
- PET can also be used to examine activity within specific neurochemical pathways in the brain, so that neurochemically specific changes as a function of cognitive state can be assessed.
- SPECT is very similar to PET in the way that it works. It has generally poorer spatial and temporal resolution, but is much cheaper to use.
- NIRS is a non-invasive technique that uses light to image the cortex nearest the skull. It is particularly useful for imaging brain activity in babies and children.
- Imaging data must be extensively processed before they can be analysed. Data processing steps include realignment, spatial normalization, spatial smoothing, and temporal filtering.
- Data analysis is usually conducted independently at every point in the brain. Models that predict signal change over scans are tested against observed values. If all points in the brain are tested, significance thresholds must be adjusted to control for Type I errors (or false positives). Alternatively, researchers can restrict their analyses to the part of the brain they are interested in, thereby reducing the multiple-comparisons problem and permitting use of lower statistical thresholds.

4 TRANSCRANIAL MAGNETIC STIMULATION (TMS)

In TMS, a brief current is induced in brain tissue through the scalp and skull, using a brief magnetic pulse delivered via a coil (about the size of a hairdryer) placed over the head (see Figure 14.2(g)). The magnetic field, which is quite strong although very brief, passes through the scalp and skull and into the brain.

The shape of the coil determines the properties and size of the field: a **figure-8 coil**, which produces a very strong and focused field at the brain location that is deep relative to the crossing in the '8', is the most popular. The induced current in the brain then renders the local tissue more or less active, depending on the TMS technique that is employed. TMS is non-invasive, is apparently free of serious side-effects (if used properly; see BioMag Laboratory, 2008: http://www.biomag.hus.fi/tms/safety.html), and has considerable potential as a research, diagnostic, and even therapeutic tool.

As a research tool for exploring the mapping of cognitive functions in the brain, TMS is used to temporarily and reversibly disrupt normal brain function in healthy subjects. This approach is conceptually similar to classic neuropsychology, in which patients with focal brain damage are carefully assessed for evidence of cognitive deficits. Any impaired cognitive function is assumed to relate to the damaged bit of brain. In TMS, activity in an area thought to be necessary for the performance of a particular function can be temporarily interrupted, and then performance on tasks relevant to that function can be assessed. Normal function resumes within a few minutes after actual stimulation has stopped, and so tests are usually administered at the same time as the TMS pulses, or shortly thereafter.

Because the field is created on the surface of the head, there is a trade-off between the size of the brain area that is stimulated and the depth to which the stimulation can reach. Even the strongest and most focused fields cannot penetrate very much further than the surface of the cortex, so that only surface locations (including motor cortex and other frontal areas) are suitable for study. Also, the temporalis muscle, which lies between the scalp and the skull over the temporal lobe, can be caused to contract quite painfully with TMS stimulation, and so TMS is not often used over the temporal lobe.

TMS, like neuropsychology, provides information about what parts of the brain are critical for particular cognitive functions. This is unlike fMRI and PET, which cannot differentiate between areas showing activity that is merely correlated with a particular cognitive function, and areas that are essential to the function.

TMS is used diagnostically to evaluate the patency of connections between the brain and the muscles in disorders like stroke and multiple sclerosis. TMS also appears to be a useful treatment of a number of psychiatric disorders, including depression, anxiety, and some of the symptoms of schizophrenia (Slotema *et al.*, 2010).

SUMMARY OF SECTION 4

- TMS can be used to disrupt brain function temporarily. When used properly, it is a safe way to induce a 'reversible lesion' in a particular region of the brain, to determine whether that region is critically involved in the performance of a given task.
- It is only suitable for use with regions on the surface of the brain, and cannot easily be used to study function in the temporal lobe.

5 CHOOSING A TECHNIQUE

In general, choosing which neuroimaging technique to use for a particular purpose requires researchers to consider a number of different factors. Try Activity 14.1.

✱ ACTIVITY 14.1

There is a theory that states that speech perception is not purely auditory. The hypothesis is that when people hear speech, it automatically activates motor regions involved in pronouncing speech sounds. Figure 14.18 shows where motor areas related to speech production are likely to be. Try to think of ways in which you could use two different techniques you have learned about to evaluate this theory. Note down the reasons for your choice of technique (why you chose the ones you did, and why you didn't choose the others). How might you set up an experiment to evaluate the theory? What kind of volunteers would you recruit? What kind of experimental design would you use? What would you measure? How would you establish whether you had a significant effect? Note down your answers to these questions. Would the results reveal what areas might be involved in speech perception versus those absolutely necessary for speech perception?

COMMENT

The most appropriate techniques, given the question, are TMS and fMRI. MEG/EEG aren't useful since the question is about where in the brain the activation is observed, and fine temporal resolution isn't really necessary. PET and SPECT both have a radiation load, and since this experiment can be performed in normal volunteers (without metallic implants), fMRI is most suitable. An argument could be made that PET is better since this would necessarily be an auditory study and fMRI is noisy, but fMRI can be used for auditory studies as long as they are designed properly. Specifically, the fMRI study should use intermittent (sparse) imaging, where stimulus presentation is alternated with scan acquisition. Experimental design could either be parametric (possibly varying rate or amount of speech) or subtractive. If subtractive, then the experimental conditions should involve listening to speech (isolated sounds, single words, or sentences). Control conditions should include one that

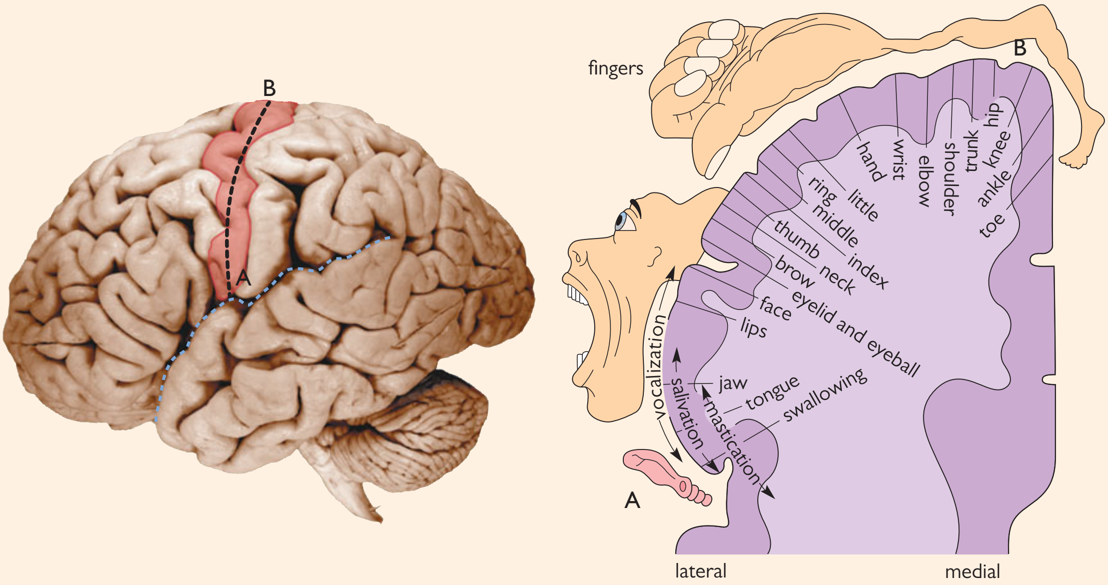


FIGURE 14.18 The 'motor homunculus'. The primary motor region (A–B) represents different parts of the body in an orderly way. The part of the motor cortex involved in mouth and tongue movements (and so speech production) is found at the lower end, near the Sylvian fissure (blue dotted line).

uses non-speech sounds matched to the speech sounds on duration (as in the hierarchical subtraction fMRI example, Figure 14.11). The analysis would either look for regions that correlated with the parametric factor (in the case of a parametric design) or regions in which activation during experimental conditions was higher than during control conditions. Since a region of interest is included in the hypothesis, the problem of multiple comparisons can be lessened, and the significance threshold used can be more liberal than if tests were conducted over the whole brain.

A TMS experiment would involve delivering TMS over the inferior motor region while the volunteer hears speech, and then testing their perception or comprehension of the stimulus. This could be compared to TMS over other regions of the brain (not expected to be involved in speech perception) while the volunteer hears speech. Any stimulus recall, recognition, or comprehension test, administered after the speech stimulus (and TMS pulse), could be the dependent measure. Changes in performance on the task would be examined as a function of brain area stimulated: if the inferior motor region is necessary for speech perception, performance will be impaired after TMS stimulation there. If the motor region is either involved but not critical or is not involved at all, task performance will not change as a function of stimulation site.

6 SUMMARY AND CONCLUSIONS

This chapter has introduced you to several methods for the study of normal brain function in healthy volunteers. These are summarized in Table 14.1. Most of these techniques are also popular for use with patients, to examine brain function in people with specific diseases or mental disorders. EEG and MEG give excellent temporal resolution and are sensitive to neuronal activity directly, although they reflect aggregate activity over large groups of neurons. It is hard, without knowing exactly how many groups of neurons are active, to locate EEG and MEG activity with any precision. fMRI, PET, SPECT, and NIRS are sensitive to local changes in brain metabolism, particularly blood flow and the oxygenation state of blood. These are indirectly related to neuronal activity and happen much more slowly, so temporal resolution is not very good. However, these metabolic changes happen in or very close to active neural tissue, and so sources of activation can be located with confidence (especially for fMRI). PET and SPECT can also be used to study the relationship between specific neurotransmitter systems and cognitive function. Finally, TMS can be used to temporarily inactivate neural tissue close to the skull, in order to determine the importance of that region of the brain for particular cognitive functions.

Each of these techniques has its unique advantages and disadvantages. As technology progresses, we can expect the temporal and spatial resolution of all of these tools to improve somewhat, although we are near the practical limits of what is possible already. A major challenge for the coming years will be to combine data from these different methods in order to overcome the limitations of any individual technique. For example, combining EEG and fMRI would give us the excellent temporal resolution of EEG combined with the excellent spatial resolution of fMRI. Combining data across these tools will provide pictures of the functional organization of the brain that are unprecedented in their detail and richness.

TABLE 14.1 A summary of methods used to study the relationship between brain and behaviour in normal volunteers

Technique	Spatial resolution	Temporal resolution	Physiological basis
EEG	>10 mm*	1 ms or less	Summed electrical activity over large groups of neurons
MEG	>5 mm*	1 ms or less	Summed magnetic fields produced by large groups of active neurons
fMRI	<5 mm	200 ms–1 sec ‡	Magnetic field fluctuations produced by local changes in the proportions of oxyhaemoglobin and deoxyhaemoglobin in blood
PET	5–20 mm†	10 sec– minutes†	Radioactive decay occurring where a labelled compound (water in blood, neurochemical) has accumulated
SPECT	20 mm	minutes	Radioactive decay occurring where a labelled compound (neurochemical) has accumulated
NIRS	30 mm	100 ms	The vascular system's reflection/absorption characteristics change as a function of blood flow/oxygenation; changes in transmitted near-infrared light and received light are measured
TMS	5–10 mm	<0.5 sec	Strong focal magnetic field transiently disrupts activity in large groups of neurons

** A general reliable figure cannot be given, since for both EEG and MEG spatial resolution depends on several factors such as modelling assumptions, number of sensors, or signal-to-noise ratio.*

† Spatial and temporal resolution in PET depend on the compound that is used. Compounds that act specifically within particular neurotransmitter systems can take longer to accumulate detectable signal but can be localized with greater precision (since we know where these neurotransmitter systems are found) than labelled water.

‡ The lower bound of this range refers specifically to when the timings of two events are measured relative to each other within the same brain area of the same subject.

FURTHER READING

Hansen, P., Kringelbach, M., and Salmelin, R. (eds.) (2010) *MEG: An Introduction to Methods*, New York, Oxford University Press. Neuroimaging is a fast-moving and relatively young discipline and, as a consequence, it is difficult to find accessible textbook treatments of the core issues. Accessible chapters on EEG and, in particular, MEG are rare but here at least there is a recent textbook, albeit somewhat advanced.

Other readings are suggested according to the kind of neuroimaging technique on which they focus.

Molins, A., Stufflebeam, S.M., Brown, E.N., and Hamalainen, M.S. (2008) 'Quantification of the benefit from integrating MEG and EEG data in minimum l2-norm estimation', *Neuroimage*, vol.42, no.3, pp.1069–77. This methodological study describes the benefit that can be gained by combining EEG and MEG.

Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R. Jr., *et al.* (2000) 'Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria', *Psychophysiology*, vol.37, no.2, pp.127–52. Guidelines for the analysis of ERPs. Also discusses basics of experimental design and provides an overview of standard analysis techniques. Not only useful for publishing papers, but also for their evaluation.

Naatanen, R. (2001) 'The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm)', *Psychophysiology*, vol.38, no.1, pp.1–21. Overview of one of the most common paradigms in ERP research ('passive oddball paradigm') and the associated 'mismatch negativity' (MMN) component in EEG and MEG. It describes basic concepts of ERP/ERF analysis, visualization, and data interpretation.

Sharon, D., Hamalainen, M.S., Tootell, R.B., Halgren, E., and Belliveau, J.W. (2007) 'The advantage of combining MEG and EEG: comparison to fMRI in focally stimulated visual cortex', *Neuroimage*, vol.36, no.4, pp.1225–35. Study comparing simple visually-evoked brain responses among EEG, MEG, combined EEG and MEG, and fMRI. Discusses differences between techniques, different source estimation procedures for EEG/MEG, and the benefit of combining EEG and MEG for source estimation.

Huettel, S.A., Song, A.W., and McCarthy, G. (2009) *Functional Magnetic Resonance Imaging*, 2nd edition, Sunderland, MA, Sinauer. Comprehensive textbook covering all aspects of functional magnetic resonance imaging, including experimental design, signal acquisition, data processing, and analysis.

Henson, R. (2005) 'What can functional neuroimaging tell the experimental psychologist?' *Quarterly Journal of Experimental Psychology Section A*, vol.58, no.2, pp.193 –233. A thought-provoking review of the application of fMRI to psychology.

Lloyd-Fox, S., Blasi, A., and Elwell, C.E. (2010) 'Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy', *Neuroscience and Biobehavioral Reviews*, vol.34, no.3, pp.269–84. (Epub 24 July 2009) A review of NIRS as a technique for asking questions about functional brain development.

George, M.S. (2003) 'Stimulating the brain', *Scientific American*, vol.289, no.3, pp.67–73. An accessible review of TMS.

REFERENCES

BioMag Laboratory, Helsinki University Central Hospital (2008) *Safety of TMS*, www.biomag.hus.fi/tms/safety.html (accessed 15 Feb 2011).

Churchland, P.S. and Sejnowski, T.J. (1988) 'Perspectives on cognitive neuroscience', *Science*, vol.242, pp.741–5.

Cohen, M.S. and Bookheimer, S.Y. (1994) 'Localization of brain function using magnetic resonance imaging', *Trends in Neurosciences*, vol.17, pp.268–77.

Cox, D.D. and Savoy, R.L. (2003) 'Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex', *Neuroimage*, vol.19, pp.261–70.

Donders, F. (1969) 'On the speed of mental processes', in Koster, W.G. (ed.) *Acta Psychologica, Attention and Performance II*, Amsterdam, North-Holland Publishing.

Friston, K.J., Price, C.J., Fletcher, P., Moore, C., Frackowiak, R.S., and Dolan, R.J. (1996) 'The trouble with cognitive subtraction', *Neuroimage*, vol.4, pp.97–104.

Kim, D.S., Duong, T.Q., and Kim, S.G. (2000) 'High-resolution mapping of isoorientation columns by fMRI', *Nature Neuroscience*, vol.3, pp.164–9.

Klein, D., Milner, B., Zatorre, R.J., Meyer, E., and Evans, A.C. (1995) 'The neural substrates underlying word generation: a bilingual functional-imaging study', *Proceedings of the National Academy of Sciences of the USA*, vol.92, pp.2899–903.

Koepp, M.J., Gunn, R.N., Lawrence, A.D., Cunningham, V.J., Dagher, A., Jones, T., Brooks, D.J., Bench, C.J., and Grasby, P.M. (1998) 'Evidence for striatal dopamine release during a video game', *Nature*, vol.393, pp.266–8.

Kriegeskorte, N., Mur, M., Ruff, D., Kiani, R., Bodurka, J., Esteky, H., *et al.* (2008) 'Matching categorical object representations in inferior temporal cortex of man and monkey', *Neuron*, vol.60, no.6, pp.1126–41.

McArthur, C., Jampana, R., Patterson, J., and Hadley, D. (2011) 'Applications of cerebral SPECT', *Clinical Radiology*, vol.66, no.7, pp.651–61.

Menon, R.S. and Kim, S.G. (1999) 'Spatial and temporal limits in cognitive neuroimaging with fMR', *Trends in Cognitive Sciences*, vol.3, pp.207–16.

Mur, M., Bandettini, P.A., and Kriegeskorte, N. (2009) 'Revealing representational content with pattern-information fMRI: an introductory guide', *Social Cognitive and Affective Neuroscience*, vol.4, no.1, pp.101–9. (Epub 17 Jan 2009)

Opitz, B., Mecklinger, A., Von Cramon, D.Y., and Kruggel, F. (1999) 'Combining electrophysiological and hemodynamic measures of the auditory oddball', *Psychophysiology*, vol.36, pp.142–7.

Raichle, M.E. (1998) 'Behind the scenes of functional brain imaging: a historical and physiological perspective', *Proceedings of the National Academy of Sciences of the USA*, vol.95, pp.765–72.

Sharon, D., Hamalainen, M.S., Tootell, R.B., Halgren, E., and Belliveau, J.W. (2007) 'The advantage of combining MEG and EEG: comparison to fMRI in focally stimulated visual cortex', *Neuroimage*, vol.36, no.4, 1225–35.

Slotema, C.W., Blom, J.D., Hoek, H.W., and Sommer, I.E.C. (2010) 'Should we expand the toolbox of psychiatric treatment methods to include repetitive transcranial magnetic stimulation (rTMS)? A meta-analysis of the efficacy of rTMS in psychiatric disorders', *Journal of Clinical Psychiatry*, vol.71, no.7, pp.873–84.

Talairach, J. and Tournoux, P. (1988) *Co-planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System: An Approach to Cerebral Imaging*, New York, Thieme.

Tallon-Baudry, C. and Bertrand, O. (1999) 'Oscillatory gamma activity in humans and its role in object representation', *Trends in Cognitive Science*, vol.3, no.4, pp.151–62.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

CHAPTER 15

COGNITIVE NEUROPSYCHIATRY

David Okai, Angus Gellatly, and Anthony David

1 INTRODUCTION

Cognitive neuropsychiatry is, in relative terms, a new discipline that attempts to draw upon the success of its sister disciplines, cognitive psychology and cognitive neuropsychology, to explain the symptoms of (neuro-) psychiatric disorders, such as dementia or schizophrenia. It represents a systematic and empirically driven approach to explaining clinical psychopathologies in terms of deficits in normal cognitive processes, and illustrates a rapprochement between cognitive neuropsychology, clinical medicine, and the neurosciences in addressing common questions about disorders of mind and brain.

There is a reciprocal relationship between the neurological components of medicine and psychology. Neuropsychology seeks to determine the relevance of brain damage or disorder to changes in behaviour. Conversely, the neuropsychiatrist studies manifestations of disturbed behaviour or cognitive deficits in order to aid in diagnosis.

What marks the cognitive neuropsychiatry approach as different from the rest of the medical specialties is the integration of discrete, or focal, brain pathology with assessment and understanding of pre-illness psychosocial history, allowing for a formulation of illness based on cognitive and personal understanding of the individual. See, for example, Box 15.1.

Patients seen by clinicians tend to fall into three main types: (1) those who suffer with a neurological disorder and have cognitive, behavioural, or psychiatric symptoms; (2) those who have medically unexplained neurological symptoms; and (3) those who have a mental illness with a possible organic aetiology.

Cognitive neuropsychiatry adopts a theory-driven approach, using patterns of impaired and intact cognitive performance observed in patients to revise current models of normal cognitive functioning. By definition and necessity its scope is broad. It encompasses disorders such as schizophrenia and dementia due to its phenomenological interest in symptoms such as delusions, hallucinations, and misidentification phenomena, although it maintains a focus on symptoms (and occasionally symptom clusters) rather

BOX 15.1 CASE STUDY Mr E

At the time of assessment, Mr E was a 56-year-old left-handed man. He was born in northern Italy and moved to southern Africa aged 11. He was an electronic engineer and a successful businessman. He was recently remarried and had a young daughter. His wife described him as a highly active and knowledgeable person, one who likes 'to chat to people and have control over situations'. He had no psychiatric history.

Mr E was referred to the neuropsychiatrist following an operation for a meningioma. On waking from surgery, Mr E was severely disoriented and confused, a condition that continued for the next month. Thereafter, there was a gradual improvement until he reached a stage where he was able to cope relatively independently at home. However, his memory abilities remained markedly and disproportionately impaired, and he was diagnosed with a Korsakoff-like amnesia. He was often not aware that he was married, nor did he know the names of his relatives, and his orientation in time and place was poor. A gross reduction in motivation was also reported. His expressive output was confabulatory and incoherent, and he lacked any insight into his memory deficit.

Mr E's confabulations often seemed to contain real events from his past misplaced in time or location. However, he also sometimes produced 'memories' with no relation to the realities of his life. In different sessions he claimed to own a Maserati and a Porsche, that he was working with the police doing chemical analysis, that he sold bags at Grand Prix events, and that he had two sons, the older of whom owned a motorcycle. His wife and brother reported that all these 'memories' were without foundation, and in reality his only child was his daughter.

Where true autobiographical memories are simply displaced in time, they are known as 'momentary confabulations'. Where, as in the above examples, stories are wholly without foundation, they are known as 'fantastic' or 'severe' confabulations, and these tend to be associated more with severe memory deficits.

Mr E's confabulations were typically accompanied by denial that he suffered from an unusual memory deficit (anosognosia). Although aware of some of the consequences of his poor memory, he attributed these to other factors, such as his own unwillingness to try to remember things or his age. For example:

(1) Examiner: 'Your memory is failing you?'

Mr E: 'It is. More than failing is that I am not interested in helping it along.'

(2) Examiner: 'And how about your memory?'

Mr E: 'Old age. Everyone would like to have a better [memory], but we get to a certain age in life, which you have to accept.'

Adapted from Fotopoulou *et al.* (2004).

than specific diagnoses. As such it moves beyond diagnosis and a classification system towards a method that applies a cognitive analysis of symptoms regardless of whether they are due to recognized 'discrete' or generalized brain pathology.

It is clear that a complex interplay of neural systems underlies most psychological processes, but cognitive neuropsychiatry attempts to draw from neuropsychological accounts of how the brain is thought to 'work' (see, for example, Chapter 13) and extend them to include belief formation, willed action, the concept of self, and the concept of others. The aim of this chapter is to illustrate the convergence between cognitive neuroscience and clinical psychiatry and to demonstrate how this has already begun to add to our understanding of psychiatric conditions.

We first look at the effects of lesions in the various major sub-divisions of the brain. Then, the following section examines the 'dementias', as they can serve as a useful heuristic tool to illustrate neurobiological, neuroanatomical, and cognitive neuropsychological contributions to understanding clinical manifestations of illness. Theoretical assumptions relating to psychosis (e.g. schizophrenia), affective disorders (e.g. depression), disinhibition (e.g. obsessive-compulsive disorder), impulse control disorders, and so-called conversion disorders will follow. This chapter is not intended to be an exhaustive review of all the conditions relevant to neuropsychiatry (for that, see suggestions in the 'Further reading' section).

2 FOCAL AND DIFFUSE BRAIN INJURY

As you saw in Chapter 13, there has long been interest in associations between brain region and brain function (Broca, 1861). On the whole, these associations have been inferred from observations of neuropsychological deficit in cases of relatively circumscribed brain lesion, where the organic pathology was very clear. Whilst rare, these conditions lead to conclusions about the specificity of function of certain brain regions. As you also saw in Chapter 13, caution needs to be applied to such inferences, not least because pure focal pathology is rare. Furthermore, it is unclear if the neuropsychological manifestation of any given injury is a result of discrete damage to a localized brain region or of disruption of intercommunicating pathways, where the combined function of differing brain regions is necessary for a specific purpose (Alexander and Crutcher, 1990). What is known of localizing of neurological function is described in the rest of this section. This will lead the way to a subsequent discussion of more general deficits in the forms of the various dementia syndromes in the following section.

2.1 Frontal lobe

The frontal lobes and their pathology represent a challenge to the understanding of neuropsychological and psychiatric dysfunction and treatment. From an evolutionary perspective, development of the frontal lobes is thought to separate us from other mammals and is central to 'what makes us human'. Loss of function within the frontal lobes leads to impairment in human social interaction with a release of innate drives that impact on the sufferer's social functioning. Manifestations of loss of function include problems in reasoning, judgement, self-awareness, attention, and understanding of interpersonal interactions. Specific cognitive deficits may be identified, such as problems with working memory (see Chapter 9), but there appears to be a disproportionate impairment for patients in social and occupational function. The frontal lobes have therefore been thought to be the location of so-called 'executive functions'. The use of the term 'executive', by analogy with the 'chief executive' of a business, serves to illustrate the need for the coordinated and controlled operation of various processes and sub-processes required to deal with day-to-day goals in a flexible manner (Elliott, 2003).

Marked changes in personality are also associated with frontal lobe damage and appear to comprise two main patterns, although these often co-occur (Blumer and Benson, 1975). Lesions with predominant involvement of the orbitofrontal area produce what has been termed a 'pseudopsychopathic' change, involving antisocial behaviour with lack of insight, disinhibition, euphoria, and facetious humour. By contrast, lesions involving the dorsolateral region are described as 'pseudodepressive' in nature and include symptoms such as apathy, loss of initiative, and slowing of thought and motor action. Psychometric tests have been devised to test for frontal lobe damage and these include tests of verbal fluency, the Wisconsin card sort, the Stroop test, the Tower of London task, and Trail making tests. However, it is not uncommon for a patient who displays marked disturbance in everyday functioning and symptoms clinically consistent with a frontal lobe syndrome to do relatively well on such psychometric tests (Shallice and Burgess, 1991). Hodges (2001) has suggested that this is because these tests are most sensitive to dorsolateral dysfunction, whilst many of the behaviours that bring the patient to clinical attention are orbitofrontal in origin. Similarly, in practice, diffuse damage and generalized impairment on psychological tests may coexist with a relatively defined cluster of antisocial and other behaviours.

Additionally, performance on tests of intellect and cognition can be relatively preserved but patients may need prompting to answer questions, or may persist with a test only when being overseen, or may perform better when given a structured environment (Hebb, 1949). The relatively normal performance on such tests when suitably administered may indicate that many areas of the brain are involved in contributing to psychological faculties such as intelligence and judgement, with coordination of these processes by the frontal lobes required to apply them in a meaningful fashion in day-to-day life. Such observations have led researchers to seek more environmentally meaningful tests of frontal lobe function, such as the 'Behavioural Assessment of the Dysexecutive Syndrome scale' (Wilson *et al.*, 1997) and the 'Frontal Systems Behaviour Scale' (Grace *et al.*, 2001), which are said to have more 'ecological validity' than standard psychometric tests. Finally, the role of confabulation can potentially add to interpersonal disability in this range of conditions (Chapter 13).

Spontaneous confabulation secondary to sub-cortical pathology can be perceived by those who know the patient as an attempt to provide a 'post hoc' rationale or justification of antisocial 'inexcusable' behaviour (Kopelman, 2002). In reality the patient may have very little recollection of the behaviours, which have been performed with a disregard for long-term consequence, something that is characteristic of this condition. Kopelman has described these inaccuracies or additions as a 'normal' response to faulty memory – akin to the filling-in of a story following a delay (Kopelman, 2002).

2.2 Parietal lobe

Neurological dysfunction can result in a disturbed sense of awareness on the opposite (contralateral) side of the body. This is not a pure loss of sensation, however, but reflects more a higher failure of processing of sensory stimulation. Examples include difficulty with two-point discrimination (being able to say whether one or two points on a region of the body have been touched at a time), failure to recognize objects placed in the hand, and inability to recognize numbers or letters inscribed on the palm. A visual 'sensory inattention' may be present, meaning the patient may attend to half of a plate of food or draw numbers corresponding to half of a clock face only (as in neglect; see Chapters 2 and 13). Here it is important to be aware of how the functions of the two hemispheres of the brain differ. For example, the term 'dominance' is employed to indicate which side of the brain is responsible (or dominant) for the ability to speak, as well as the capacity to understand the spoken and written word. Proportions of dominance for each cerebral hemisphere can be revealed using the Wada Test, which involves anaesthetizing one or other of the hemispheres. In 90 per cent of the population, the left cerebral hemisphere is dominant and the right hemisphere non-dominant for this kind of language processing.

Following a lesion to the non-dominant parietal lobe, there may be cognitive disturbances that affect language and number comprehension. There may be inattention to the opposite side of the body and the patient may deny that parts of the side in question (such as an arm) belong to them. This denial of disease (anosognosia) following parietal damage is in contrast to the lack of insight seen in schizophrenia, which is thought to be more frontal in origin.

Lesions of the dominant parietal lobe may be associated with difficulties in speaking (dysphasia), as well as the so-called Gerstmann's syndrome, which consists of right-left disorientation, difficulty identifying which finger is being touched on examination (finger agnosia), dyscalculia (specific difficulty with calculation and manipulation of numbers), and dysgraphia (specific difficulty with writing and the production of coherent sentences but also partly contributed to by fine motor skill loss). Finally there may be topographical disorientation (often manifesting in day-to-day life with patients being found 'wandering' as a result, or getting lost on previously familiar streets while trying to find their way home) and dressing dyspraxia (the patient may put their clothes on the wrong way round or inside out).

2.3 Temporal lobes

Dysfunction associated with temporal lobe damage can be divided into four main types: defects in auditory, visual, and memory function, and change in personality. More rarely there may be an association with a psychotic presentation.

Lesions of the dominant temporal lobe, especially in the anterior region (known as Wernicke's area), result in what is now termed a fluent dysphasia with rapid, nonsensical speech production. This may resemble the suite of language problems described in patients with schizophrenia as thought disorder (see also Section 4.2). Non-dominant side lesions have on occasion been shown to impair a patient's perception of music, faces, and drawings. Unilateral lesions usually have little if any effect on memory function, presumably due to the other hemisphere compensating for the loss. However, dominant side lesions may impact on verbal memory, and non-dominant side lesions on non-verbal (spatial) memory. Bilateral lesions result in a global amnesia, usually with preservation of working memory. Personality changes that occur are often those of emotional lability, aggression, and change (either a decrease or an increase) in sexual activity.

Three types of pathology bear special mention in relation to the temporal lobes. Firstly, Alzheimer's disease involves atrophy of the medial temporal lobe (specifically the hippocampal region and related structures). This is one of the first brain regions to be affected in Alzheimer's disease and in mild cognitive impairment (Braak and Braak, 1997; Müller *et al.*, 2005). Neuroimaging and behavioural studies suggest a critical role for the hippocampus in encoding and recall of episodic memories, with damage resulting in a reduced ability to acquire new episodic memory (anterograde amnesia). There is additionally progressive loss of memories prior to the damage (retrograde amnesia). Recent memories are felt to be affected more than distant memories, which may reflect a progressive dissociation from the hippocampus as memory becomes increasingly consolidated (Piefke *et al.*, 2003). Secondly, temporal lobe epilepsy may produce a psychosis in between epileptic attacks that is indistinguishable from the pattern of symptoms seen in schizophrenia (Toone, 1991). Finally, in schizophrenia, lesions in the temporal lobe have been identified in the planum temporale and Heschl's gyrus, areas that play a role in language and auditory processing. Some researchers have gone on to suggest that hallucinatory and cognitive deficits are associated with such abnormalities (Shenton *et al.*, 2001).

2.4 Occipital lobe

Deficits associated with occipital lobe damage may be in the form of visual impairment, such as cortical blindness. Very rarely, there is associated denial of this blindness (Anton's syndrome). The patient often voices a confabulatory account of (unlikely) reasons to explain their current visual impairment consistent with an anosognosia. The converse is the phenomenon of blindsight (see Chapter 13), where a patient who is phenomenally blind is able to process visual information that is not consciously represented (i.e. unconscious vision). With the majority of occipital lesions, however, sight is preserved but there may be disturbances of high level perception (visual cognition). One such condition is alexia without agraphia, where there is a preservation of letter recognition but whole words (including words written by the sufferer) cannot be read. Another example is where the individual sees non-complex visual hallucinations, for example of geometrical shapes (ffytche *et al.*, 1998).

2.5 Brain stem and cerebellum

Lesions here produce symptoms akin to frontal lobe impairment, due to the strong frontal projections from these structures. Problems with initiation, planning, and mental set shifting (adapting one's preparatory adjustment for performing a task when the task changes) are particularly common. The hallmark disease of this condition is Parkinson's disease (PD), which has many associated neuropsychiatric co-morbidities. Additionally Korsakoff's amnesia, with a characteristic neuropsychological profile of intact working memory but pronounced anterograde and retrograde amnesia, is a condition with brain-stem involvement. A 'cerebellar affective syndrome' has been described (Schmahmann and Sherman, 1998), although it is difficult to exclude additional cerebral lesions, especially when the cause of cerebellar damage may affect other parts of the brain, as is commonly the case in, for example, stroke. Finally, damage to these structures can affect intelligence or a patient's level of arousal, or may produce appetite disturbance (such as overeating) and unpredictable swings in emotion (Trimble and Cummings, 1981).

SUMMARY OF SECTION 2

- Frontal lobe dysfunction is associated with disorders of conduct, with lesions being described as 'pseudodepressive' or 'pseudopsychopathic' in nature.
- Parietal lobe dysfunction can result in a disturbed sense of awareness, inattention, or language and calculation problems.
- The temporal lobes are associated with dysphasia, memory impairment (especially if there is bilateral temporal lobe damage), and personality change.
- Occipital lobes are associated with visual impairment including blindness and visual hallucinations. The hallucinations are associated with varying degrees of insight.
- Brain stem pathology mirrors that of frontal lobe pathology due to strong interconnections. It may also be associated with impaired control of actions and emotions.

3 DEMENTIA

Dementia is a non-aetiologically specific term applied to a range of neurodegenerative disorders that involve progressive loss of higher intellectual functioning (see DSM-IV of the American Psychiatric Association, 2000). The deficits must represent a decline from a previous level of functioning in order to meet the definition of the term dementia, and they must be of such severity as to impact on the sufferer's work or social functioning. A central component is amnesia or memory impairment, with individuals becoming impaired in their ability to learn new information or showing greater forgetting of previously learned information. The memory disorder must be accompanied by at least one other symptom from the list in Table 15.1. Factors such as personality change and social conduct are also important in diagnosis.

Neurodegeneration is the main reason for progressive deterioration in neuropsychological function in dementia and may result in generalized presentation of symptoms, with a clinical profile characteristic of many of the deficits described in Section 2, due to many brain regions being affected. The most common example of this is Alzheimer's dementia. Alternatively, a patient may present with a relatively specific cognitive profile that ultimately becomes more generalized as the condition progresses (e.g. Parkinson's disease).

TABLE 15.1 Clinical profile of deficits in dementia

Deficit	Cognitive dysfunction
Aphasia	Deterioration in language function
Apraxia	Impaired ability to perform motor activities despite intact motor ability
Agnosia	Failure to recognize or identify objects despite preserved sensory function

A useful clinical division resides in differentiation between the dementia syndromes that mainly affect the '**cerebral**' cortex – primarily temporal, parietal, and occipital lobes, which as we have seen result mainly in aphasia, apraxia, and agnosias – and those that impact mostly on '**sub-cortical**' structures (the area of the brain immediately below the cortex), with a predominant effect on the functioning of pre-frontal areas and the basal ganglia. The clinical manifestations of these latter sub-cortical forms of dementia are executive dysfunction, slowed motor and processing speed with relative inflexibility of thought, and personality change. The following sections will outline the neuropsychological presentation of specific dementia syndromes.

3.1 Alzheimer's disease

Alzheimer's disease (AD) is a cortical dementia and the most common neurodegenerative disease. It affects 4 per cent of those aged over 65 and whilst age is the most important risk factor, the disorder is not simply a condition of accelerated ageing. Common neuropsychiatric consequences include depressive symptoms, paranoid delusions (particularly of theft), visual hallucinations, and personality change. The aetiology is unknown but likely to be multi-factorial, with post-mortem studies indicating a direct correlation between extent of brain pathology and extent of cognitive impairment.

A unique longitudinal study by Snowdon and colleagues (n = 678) examined nuns, who were members of a single congregation and aged between 75 and 102 at initial study contact (Snowdon *et al.*, 1996). Neuropsychometric tests were performed annually thereafter and, importantly, all the nuns agreed to brain donation at point of death. The study allowed examination of a wide range of dementing illnesses – from severe Alzheimer's and vascular dementia, through to participants with little or no cognitive impairment. Clinical assessment included tests such as the Mini-Mental State Examination (MMSE) and delayed word recall (a test of short-term memory from a list of ten words). However, it is important to recognize that core to a diagnosis of dementia is impairment in activities of daily living (for instance, using the telephone or correct calculation of change in shopping activities). These were also assessed. Clinical findings could then be compared with neuropathology and included a positive correlation between the severity of dementia and the degree of neuropathology present in the brain. On the whole, however, a sizeable proportion of participants with high levels of Alzheimer's pathology appeared free of clinical expression of symptoms, including 8 per cent of those with the most severe pathology. The process of resistance to disease appeared mediated by higher levels of education and intelligence, indicating that this group had higher levels of cognitive reserve (namely the brain's, or the individual's, ability to adapt to or compensate for the presence of pathology).

This study had several strengths. Firstly, because of the brain donations it allowed for histological confirmation of clinical diagnosis (i.e. through study of cell structures). The majority of studies employ criteria that allow at best a 'probable' diagnosis of a specific dementia (e.g. Alzheimer's, or vascular, or fronto-temporal), as histological confirmation is usually not possible. Additionally, the study allowed for elimination of many factors that confound epidemiological research, due to the relative homogeneity of the sisters' environment and lifestyle. Unfortunately, this to some extent also limits the generalizability of the findings. Nevertheless, the wide range of function and health among the participants allowed a unique opportunity to understand the biological and social substrates of illness and health. It currently remains unclear how upbringing or environmental factors can contribute to a process of resistance to dementia, warranting further research on the subject.

Some causes of earlier onset of Alzheimer's disease have also been identified, which are rare but have a clear-cut genetic link, being passed down from parent to child. These tend to have higher levels of demonstrable neuropathology (scoring more highly on neuropathological scales of severity of illness) and clinically are rapid in onset with more aphasia and fronto-temporal distribution. Importantly, by far the majority of cases of Alzheimer's are multi-factorial, and therefore there is no indication that, should one family member develop the disorder, a sibling or child will also develop the condition.

3.2 Vascular dementia

Vascular dementia (VascD) is a dementia syndrome cause by impaired blood supply (and therefore reduced oxygen) to the brain. It is the second most common form of dementia. It typically results from the occurrence of multiple strokes of varying sizes over time, causing a relatively abrupt onset of dementia with a progressive, stepwise decline. In the majority of cases, neuropathology consists of recurrent lesions in the sub-cortical areas of the brain. However, DSM-IV requires additional evidence of cortical pathology (see above) of a severity to produce 'hard' or focal neurological signs, i.e. identifiable neurological symptoms that can be associated with a region of the brain consistent with the origin of the stroke. An example might be weakness of the right side of the body (known as hemiplegia) due to a stroke in the left frontal part of the brain known as the motor cortex.

The cognitive profile of a vascular dementia is typically idiosyncratic with patients often showing a 'patchy' distribution of deficits, in contrast to the more global picture typically seen in Alzheimer's disease. However, the divide between the two conditions is difficult to draw due to the large overlap in risk factors for both conditions (such as high blood pressure, history of smoking, and age). Indeed, Snowdon's study of nuns described earlier found an interaction in pathology, with those individuals showing more evidence of multiple strokes also being more likely to succumb to Alzheimer's pathology (Snowdon, 2003). These patients did not necessarily meet clinical criteria for vascular dementia, as despite there being demonstrable strokes on post-mortem examination, these may not necessarily result in focal clinical neurological signs. In truth, a significant proportion of patients who have MRI head investigation demonstrate changes in the brain consistent with multiple strokes but may be clinically asymptomatic. The role of the neuropsychiatrist may be to comment on whether they feel these changes are incidental or indeed related to a specific neuropsychiatric condition.

Most common neuropsychiatric features of VascD manifest as depression and emotional lability, especially in those with deficits in executive function. Staekenborg and colleagues (n = 484) compared the relative severity and frequency of disease in those who fulfilled criteria for VascD (Staekenborg *et al.*, 2009). Apathy, depression, and irritability were the most common complaints on screening/severity questionnaires, with sub-cortical disease more likely to present with apathy and cortical disease more likely to present with agitation/aggression.

3.3 Dementia with Lewy bodies (DLB)

This is the third most common form of dementia, often co-occurring with Alzheimer's disease pathology. Diagnostic criteria in addition to amnesia include two of three core symptoms (McKeith *et al.*, 2000):

1. tremor, rigidity (stiffness), and bradykinesia (slowed thought and action)
2. hallucinations
3. fluctuating confusional state.

These symptoms are also classically found in Parkinson's disease but are seen in a range of other conditions (such as in some patterns of vascular dementia where the strokes happen to be distributed in the same part of the brain that is usually responsible for causing Parkinson's disease, or following certain medications). In these occurrences they are referred to as parkinsonian symptoms. Lewy bodies themselves, from which the dementia is derived, are abnormal, insoluble complexes formed within neurons when attempting to eliminate damaged protein structures from within the cell. They are found in both the brain stem and the cerebral cortex and do not occur in normal ageing. The pathology is said to be present in both Alzheimer's disease and DLB. Historically, a distinction has been made between DLB and Parkinson's disease, with the suggestion that a classification of DLB should be made if amnesia occurs within a year of the onset of Parkinsonism, whilst a classification of Parkinson's dementia is presumed if amnesia occurs after this time period. Recent work has failed to support this arbitrary cut-off, instead suggesting that DLB and Parkinson's dementia lie along a spectrum of disease pathology (Ballard *et al.*, 2006). The cognitive profile is sub-cortical, with neuropsychiatric conditions including delusions and depression, as well as characteristic well-formed visual hallucinations (e.g. the patient having the sensation of perceiving a child or perhaps an animal), which are more likely to persist than in Alzheimer's dementia (Ala *et al.*, 1997).

3.4 Fronto-temporal dementia

Fronto-temporal dementia (FTD) is a collective term for a range of dementias associated with varying underlying neuropathological findings. As such it is purely a clinical diagnosis, although its presence may warrant further investigation relating to contributing pathology.

Three main types have been described: frontal variant FTD (fvFTD); progressive non-fluent aphasia; and semantic dementia. fvFTD presents with a syndrome most reminiscent of the dorsolateral conditions described above (see Section 2.1), with disinhibition and repetitive and/or antisocial behaviour. Memory is relatively spared but there is poor performance on tests of executive function. There may be a progressive dysphasia (difficulty with speech) marked by decline in language, with disintegration of grammar but usually preservation of comprehension in the early to middle stages of the disease. Semantic dementia is the term used when the temporal lobe component to this condition is predominant. Patients present with preservation of episodic memory but loss of memory for words, with marked deficits in naming of defined lists, and additional loss of general knowledge. Speech remains fluent, although the patient may make paraphrasic errors (e.g. saying dog instead of cat, or clock instead of watch) or use circumlocutions to compensate for missing words. There may also be a reversal of the usual temporal gradient, with preservation of recent compared with distant memories.

3.5 Mild cognitive impairment

Much has been made of the impact of impairment on social or occupational functioning in the definition of and criteria for dementia. What then of 'subsyndromal' forms of any of the conditions above?

For example, a barrister may score just above the threshold for diagnosis on a dementia screening tool such as the CAMCOG (a brief neuropsychological battery designed to assess the range of cognitive functions required for a diagnosis of dementia and to detect mild degrees of cognitive impairment). However, the same individual may be otherwise unimpaired in their activities of daily living. This case example would not meet the criteria for dementia; however, given the evidence of apparent decline in cognitive function, the neuropsychiatrist may be inclined to re-refer the patient for neuropsychological assessment in a year's time to see if there is any change in their cognitive score. The neuropsychiatrist may also prescribe medication in order to decrease the rate of cognitive decline, which may slow the progress of the disease on average by 6 to 12 months (Burns *et al.*, 2000).

Conditions such as this are increasingly adopted into the term mild cognitive impairment (MCI) (Petersen *et al.*, 2001). Research focus on this pre-disease state is increasing, with the hope that it may lead to better predictions of who will go on to develop full-blown dementia, study of biological disease markers that may affect such a development, and prospective treatments that may prevent conversion (i.e. treatments given in the early stages of MCI before dementia development). Current conversion rates from MCI to dementia are estimated at somewhere between 2 and 30 per cent per year (Bruscoli and Lovestone, 2004). Variation in the estimate is likely due to inconsistencies and variability in the criteria used in selection of the populations that have been studied. It has been suggested that a 'critical mass' of pathology has to be reached before prodrome (early symptoms not sufficient to meet diagnosis) becomes condition (e.g. Alzheimer's disease). However, debate still continues as to how to accurately define MCI; for example, what would constitute a deviation in cognitive or intelligence score from age-adjusted norms sufficient to warrant a diagnosis of MCI? These questions are important in that the prognosis for individuals with MCI appears to be worse than for the general population – an increased rate of mortality has been suggested (Gussekloo *et al.*, 1997).

SUMMARY OF SECTION 3

- Dementia is a syndrome involving the 'four A's', namely amnesia, aphasia, apraxia, and agnosia.
- The clinical and cognitive profiles differ for cortical and sub-cortical dementias.
- Alzheimer's disease is associated with a global cognitive decline.
- Vascular dementia is associated with a 'patchy' distribution of cognitive deficits.
- Dementia with Lewy bodies appears to lie along a spectrum of disease states with similar cognitive deficits.
- Fronto-temporal dementia presents with a relative sparing of memory, but patients may manifest language or semantic deficits.
- Mild cognitive impairment is a term used where there appears to be a decline in cognitive function that is not of a sufficient level to be classified as dementia. Increasing research is being conducted into this condition.

4 PSYCHOTIC ILLNESSES

The German psychiatrist Kraepelin popularized the concept of dementia praecox, which in time would become what many consider to be the archetypal psychotic illness, schizophrenia (Kraepelin, 1919). Whilst he would later downplay the prognosis of the condition, he maintained that at its heart was a cognitive profile that manifests as '*Verblödung*' (deterioration), '*Schwachsinn*' (mental weakness), or '*Defekt*' (defect).

Since this time more empirical work has supported Kraepelin's initial observations, namely of cognitive deficits in patients with psychosis. It is clear that there is an association between schizophrenia and neurological abnormality, with a higher frequency of neurological signs in schizophrenia patients and their relatives than in other primary psychiatric comparison groups. In contrast to the focal, or hard, neurological signs described in Section 3.2, these have been labelled 'soft signs' as they are not readily localizable to a specific part of the brain. Examples include impaired motor coordination or the abnormal preservation of neonatal reflexes commonly lost at 6 months of age, such as the tendency to grasp objects placed in the hand. It seems a logical assumption that the underlying neuropsychological deficits are equally complex.

Of further interest are the neurological conditions that present with an increased likelihood of psychosis (e.g. Parkinson's disease and epilepsy). This section will consider the neurocognitive profiles differentiating psychotic illness (usually schizophrenia) from the normal population. This will be followed by a discussion of what is known of cognitive dysfunction associated with the range of symptoms under the umbrella term psychosis. We will then consider more demonstrably organic conditions and how what is known of their pathology can aid this discussion.

4.1 Normal condition

Schizophrenia in the general population is known to be associated with cognitive deficits. David *et al.* (1997) were able to study such effects in a group of 50 000 males conscripted into the Swedish army between 1969 and 1970. Tests of verbal and visuospatial abilities, general and mechanical knowledge, and background data such as age were recorded at conscription. These tests had previously been standardized against an entire conscript cohort from a previous year. In later life, 195 of the sample subjects were admitted to hospital with schizophrenia. Low IQ, and in particular poor performance on a verbal task and the mechanical knowledge test, even following correction for general intellectual functioning, emerged as a clear risk factor for those later diagnosed with schizophrenia. This finding has been replicated in further studies, with schizophrenia groups appearing on average to have IQ scores of 4–8 points less than the normal cohort average (Reichenberg *et al.*, 2002). However, given this is the average value, the findings indicate that a substantial proportion of the general population will have IQ scores lower than those of the schizophrenia group. Similarly, not all patients with schizophrenia have low IQ.

There are two main possible implications of this that relate to our consideration of psychosis. Firstly, there may be a causal link. For example, low IQ may compromise abilities such as conceptual reasoning, resulting in psychosis (or, in other words, high IQ is protective). Secondly, there may be a confounding factor, i.e. brain abnormality may lead to both low IQ and schizophrenia. Regarding the latter, evidence indicates that obstetric-related cerebral damage does not account for the association between psychosis and intellectual functioning (Gunnell *et al.*, 2002). Additionally, there appears to be no overt association between cognitive impairment and brain structure as revealed by MRI when comparing patients with schizophrenia who are cognitively impaired with those who are unimpaired (Ortiz-Gil *et al.*, 2011).

What is known is that additional decline occurs within the first few years of psychotic illness, but not in all groups (see Table 15.2; Weickert *et al.*, 2000). Functioning then usually remains stable, with the exception of groups who are chronically hospitalised

TABLE 15.2 Cognitive impairments in patients with schizophrenia. Adapted from Weickert *et al.*, 2000

Schizophrenia group	IQ	% of total schizophrenia sample	Impaired neuropsychological functions
General intelligence decline from pre-morbid level	Initially average IQ Approx. 10 points decline in IQ	50	Executive function, memory, and attention deficits
Consistently low intelligence both pre-morbidly and subsequently	Low pre-morbid IQ No IQ decline	25	Executive function, memory, attention, and language and visual processing deficits
Intelligence in the normal range both pre-morbidly and subsequently	Average IQ No IQ decline	25	Executive function and attention deficits

where it may deteriorate further in old age (Palmer *et al.*, 2009). The majority of people with schizophrenia appear to have had satisfactory levels of functioning at school, although not necessarily at the level of their peers (or siblings). It may be that deficits of attention and executive functioning may be central features of schizophrenia independent of variations in IQ.

The cognitive deficits outlined above manifest in tasks such as the Wisconsin card sorting and Stroop tests. However, it is unsurprising that those with a condition that affects general intelligence may do worse at specific tests of cognition, in that, with worsening cognitive function, it becomes increasingly difficult to separate the contributions of specific vs general impairments. The deficits appear to apply to established as well as to drug-naive and first-episode cases of schizophrenia illness, with a relative sparing of visuo-spatial function.

Recent literature on schizophrenia has looked at the role of the COMT enzyme (responsible for dopamine breakdown; dopamine is known to have a role in the cause of the condition) in cognitive functioning. One variation of the gene responsible for the expression of this enzyme appears to have a greater effect on executive function and IQ than the other. Subjects with this genetic variation appear to perform worse on working memory tasks (Goldberg *et al.*, 2003) and make more perseverative errors on the Wisconsin card sort (Egan *et al.*, 2001). The suggestion is that abnormalities in dopamine breakdown impair the cognitive functions of the prefrontal cortex and additionally increase the risk of schizophrenia.

However, neuroimaging studies (mainly fMRI; see Chapter 14) that have examined blood flow in the frontal cortex of patients with schizophrenia in relation to working memory tasks have yielded inconsistent results. Weinberger and colleagues explain this by postulating reduced efficiency of prefrontal cortical processing in schizophrenics (Weinberger *et al.*, 2001). This causes them to have more activation than healthy controls (i.e. hyperfrontal, or showing increased blood flow in frontal regions) at low task demands, as they work harder to 'keep up'. As task demands increase, they begin to perform inadequately and have a drop in activation, becoming hypofrontal (reduced blood flow to the frontal region). However, Weinberger's explanation has yet to be consistently supported.

4.1.1 Correlation between psychotic symptoms and cognitive deficits

Schizophrenia is not a unitary construct and the heterogeneity (group differences) may contribute to the observed variability in cognitive performance tasks. Liddle and Barnes (1990) first pointed towards this heterogeneity in a group of patients with chronic schizophrenia using a battery of neuropsychological tests.

Factor analysis is a method used to group a number of variables (in this case symptoms) into clusters based on correlations or levels of agreement between the variables. These groups, or underlying dimensions, are called 'factors'. Analysis of neuropsychological test scores in patients with schizophrenia identified three factors or syndromes: reality distortion (delusions/hallucinations); disorganization (positive thought disorder, inappropriate affect, disorganized/bizarre behaviour); and psychomotor poverty (negative symptoms), with an associated neuropsychological profile for each, as shown in Table 15.3 (Liddle and Morris, 1991).

Schizophrenia is often described in terms of negative and positive symptoms, with negative symptoms being those that represent deficits in relation to normal functioning (e.g. flat affect), and positive symptoms being those that are not normally experienced (e.g. hallucinations). Subsequent work has shown that 'negative symptoms' such as social withdrawal/isolation, apathy, and poverty of speech are linked to memory deficits. However, 'positive symptoms' (hallucinations, delusions) are felt to have less of a clear link with neuropsychological functioning or impairment. We will discuss each in turn.

4.2 Hallucinations

Hallucinations are defined as perceptions that arise in the absence of an external stimulus. Studies demonstrate a tendency for patients with auditory verbal hallucinations ('hearing voices') to perceive meaningless

TABLE 15.3 Syndromes of chronic schizophrenia. From Liddle and Barnes, 1990

Symptoms in each syndrome	Impaired neuropsychological function in each syndrome
Psychomotor poverty syndrome	
Poverty of speech	Conceptual reasoning
Decreased spontaneous movement	Long-term memory
Unchanging facial expression	Object naming
Paucity of expression gesture	
Affective non-response	
Lack of vocal inflection	
Disorganization syndrome	
Inappropriate affect	Concentration
Poverty of content of speech	New learning
Tangentiality	
Derailment	
Pressure of speech	
Distractibility	
Reality distortion syndrome	
Voices speak to the patient	Figure-ground perception
Delusions of persecution	
Delusions of reference	

sounds as meaningful speech, to have problems interpreting speech in the background of noise, and to misperceive speech that has reduced phonetic clarity. Additionally, Frith (1992) has suggested that auditory hallucinations may be internally generated verbal discourse (inner speech) misidentified by the patient as arising from an external speaker. Indeed, brain imaging studies have shown that when performing tasks that require the generation of inner speech, those who have experienced auditory hallucinations fail to activate the medial temporal gyrus and supplementary motor areas. The result is that thoughts or actions are experienced as dissociated from a sense of will and are perceived as hallucinations (in the case of speech) or experienced as being under external control (the patient believes someone is controlling their body) in the case of motor activity (Frith, 1992).

Clinically, auditory hallucinations are often in the form of a voice speaking about the person (e.g. 'he's a bad person, look at him sitting there by himself'). They often manifest in conditions such as schizophrenia, and indeed make up part of its diagnostic basis (anatomical regions for this have already been suggested; see Section 2.3). By contrast, visual hallucinations appear much more commonly in organic conditions such as dementia. It is particularly common in dementia with Lewy bodies (see Section 3.3).

Brugger *et al.* (1996) described 'autoscopic' hallucinations, namely perception of oneself as an external object, which can lead to the delusion that one has a double (or doppelganger). This has been tentatively associated with occipital temporal lesions (usually stroke or neoplasm), although the condition has also been linked to Parkinson's disease sufferers and to schizophrenia.

4.3 Formal thought disorder

Formal thought disorder (FTD) is most closely tied with the disorganization cluster or syndrome. It has been likened to a fluent aphasia following lesions to Wernicke's area, and clinicians have been shown to have difficulty (when unaware of a patient's formal diagnosis) in differentiating between the two conditions. In schizophrenia, FTD is manifested by a deficit in two areas: (1) a failure of the 'pragmatics' of speech resulting in difficulty appreciating the conversational needs of the listener, which has been likened to a deficit in theory of mind; and (2) a deficit in semantic memory. See Box 15.2.

4.4 Negative symptoms

Negative symptoms that represent a decline or loss of function, such as lowered motivation, motor slowing, lack of initiative, and affective flattening, relate more consistently to cognitive abnormality (particularly deficits in executive function and memory) than do the other psychotic syndromes described above. Additionally, unlike other domains identified by Liddle and colleagues (which have no prognostic implications),

BOX 15.2 Semantic priming

Priming refers to a stimulus exposure directly influencing the response to a later stimulus. In semantic priming the 'prime' and 'target' are associated (e.g. blue–sky) and may belong to the same category and share certain features. For example, 'dog' would be a semantic prime for the target 'wolf'. Priming is hypothesized to work due to spreading activation within neural networks.

In *indirect priming* the target words are only indirectly related to the prime, usually via a mediating word, e.g. lemon–sweet (mediating word *sour*) or black–chalk (mediating word *white*). Schizophrenia patients with FTD show enhanced indirect semantic priming when compared to healthy controls (Pomarol-Clotet *et al.*, 2008). However, it is unclear whether this differentiates them from other schizophrenia patients without thought disorders.

negative symptoms demonstrate considerable stability over time. Using fMRI, Dolan and colleagues found that in both schizophrenia and depression, patients with negative symptoms, such as poverty of speech, showed reduced left prefrontal cortical neural activity (Dolan *et al.*, 1993). Negative symptoms are also often seen in neurological patients with frontostriatal dysfunction, such as Parkinson's disease; indeed, it is of note that the clinical profile of Parkinson's disease has much in common with that of the negative symptoms.

4.5 Delusions

In contrast to the negative symptoms, positive symptoms are those that show a presence of abnormal mental function such as abnormal beliefs (delusions) and abnormal perception (hallucinations). The cognitive abnormalities subserving delusions are now thought to have a direct bearing on the design of psychological interventions such as cognitive behavioural therapy. This is in contrast to an earlier view that delusions were not amenable to understanding and, by extension, treatment by talking therapies. Four main cognitive mechanisms have been identified relating to delusional pathology:

1. Jumping to conclusions: namely a specific data-gathering bias, leading patients to terminate data collection early on with insufficient weighing of evidence before coming to a strong conclusion (for example, deciding a purse has been lost if it is not in its usual place, without making any attempt to look for it in alternative places).
2. Attributional bias: deviances in attribution with a tendency to externalize, i.e. to view others rather than themselves as to blame in a predicament (see also the related discussion of attributional bias in Chapter 16).
3. Theory of mind deficits: deficits in understanding the thoughts, desires, and beliefs of others (Sprong *et al.*, 2007).
4. Bias against disconfirmatory evidence: patients appear less able to disengage from their initial interpretations of data, even when these are shown over the course of testing to become increasingly improbable (Moritz and Woodward, 2007).

Many of these concepts can be best understood in neurobiological terms by relation to a state of 'hypersalience' or 'hyperimportance'. This means more importance is given to certain aspects of the environment than is justifiable (Kapur, 2003). Dopamine in particular is felt to have a role in the salience distortion and the pathophysiology of psychosis. For example, abnormal dopamine levels may cause the patient with schizophrenia to believe the television or newspaper may have special salience or meaning. A secondary inference may then form to make some sense of this apparently special salience; for example, 'It must be because the government is trying to watch me through the television and trying to communicate with me via the newspapers'. These neurobiological processes may help researchers to better understand the cognitive processes of belief (or delusion) formation, as well as the relation between this and patients' insight into their condition (David, 1990).

What of other organic states? Delusions may be a prominent feature of conditions such as Alzheimer's, where their complexity and severity are proportional to the level of cognitive impairment. They too may result from inferential bias and in the majority of cases they are secondary to the memory deficit. An example of the most common delusion is the belief that the patient is a victim of theft (usually following the patient misplacing a household object and then blaming a relative for its misappropriation). Box 15.3 demonstrates the type of delusion that may present to a neuropsychiatrist.

4.6 Monothematic delusions: a two-factor approach

The most extensive theoretical attempt to provide a cognitive account of delusional beliefs has focused on certain monothematic delusions – that is, cases in which an individual expresses just a single delusional belief or a small set of delusional beliefs that share a single theme. Some examples of monothematic delusions are: (1) **Capgras delusion**, in which an individual, typically having suffered right hemisphere damage, believes that one or more very familiar people, such as a spouse or children, has been replaced by an identical-looking impostor; (2) somatoparaphrenic delusion, in which a person with paralysed left limbs (left hemiplegia following right hemisphere damage) claims that the limbs do not belong to them but either to someone else in the room, such as the clinician, or to someone well known to them, such as a parent

BOX 15.3 CASE STUDY Mrs A

'Mrs A', a 100-year-old Caucasian woman, was referred to the neuropsychiatric clinic by her general practitioner for evaluation of auditory hallucinations and persecutory delusions. According to Mrs A, she had begun hearing male voices some 10 months before, shortly after her 100th birthday. The voices, which she attributed to 'the Imps of Satan', sang and cursed. Mrs A was convinced that other people could hear the 'Imps' and became angry with her daughter when the latter said she could not hear them.

According to Mrs A's family, in the 3 weeks before admission, the patient became increasingly preoccupied with the 'Imps' and increasingly suspicious and fearful. On the day before admission, Mrs A frantically telephoned her granddaughter to say that people in her house were trying to kill her. In reality, the only people in the house were Mrs A's long-time home aides, whom she seemed unable to recognize. When Mrs A's granddaughter arrived at the house with her husband, Mrs A accused her granddaughter's husband of 20 years of being an impostor. Mrs A then became combative and said there was a conspiracy against her. The granddaughter called the patient's doctor, and Mrs A was brought to the Emergency Department.

Mrs A scored 27/30 on a repeat MMSE. More extensive neuropsychological testing of this patient demonstrated numerous deficits (including poor performance on the Wisconsin card sorting test and tests of verbal fluency (e.g. naming as many words as possible in a minute), as well as in multiple other cognitive domains.

Adapted from Harrington (2007) and Cervantes *et al.* (2006)

(who may not even be present); (3) mirrored-self delusion, in which an individual, again typically with right hemisphere impairment, reports that the person they see in the mirror is not them but a very similar-looking stranger.

The term 'cognitive neuropsychiatry' was actually coined by neuropsychologist Hadyn Ellis in 1991 (Coltheart, 2007) in relation to ideas about Capgras delusion first developed by Ellis and Young (1990). William James (1890/1950) had suggested that in many cases the delusions of the 'insane' were '. . . certainly theories which the patients invent . . .' to account for abnormal experiences they suffered. As described in Chapter 4, Ellis and Young (1990) hypothesized that patients with Capgras delusion may have a 'mirror image' of prosopagnosia, in that although their conscious ability to recognize faces using the ventral visual pathway is intact, they may have damage to the dorsal pathway that produces the automatic emotional arousal to familiar faces. This would lead to the experience of recognizing someone while feeling something was not 'quite right' about them because the usual emotional response to them was lacking. Attempting to make sense of this strange experience could result – in an almost rational way – in the delusion that the person is actually an impostor. Support for the Ellis and Young hypothesis came from studies showing that, as predicted, Capgras sufferers do not show differential autonomic responses to famous faces and unfamiliar faces (Ellis *et al.*, 1997; Hirstein and Ramachandran, 1997) or to pictures of family members and strangers (Brighetti *et al.*, 2007).

A somatoparaphrenic delusion could arise in a similar fashion. The patient cannot feel or move the limbs in question, therefore they cannot be theirs, and must belong either to someone else in the vicinity or to someone with similar-looking limbs, such as a parent. Along the same lines, a patient with a mirrored-self delusion was found also to suffer from 'mirror agnosia', or loss of knowledge about how mirrors work. Asked to grasp an object held up behind him to reflect in the mirror facing him, this patient reached forwards to the mirror instead of back behind him, and despite repeated attempts was unable to correct this behaviour (Coltheart, 2007). For him the mirror was not a reflecting surface but a window onto external space. Since the image of his own face appeared to be located in that external space, it would be a reasonable inference that it could not possibly *be* his own face.

A problem for explaining delusions as rational inferences (conscious or unconscious) is why the individual does not abandon their delusional belief given how unlikely it is to be true and how many other people assure them of its falsity. To make the problem even more acute, not everyone who lacks a discriminatory autonomic response to familiar and unfamiliar faces develops a Capgras delusion; the majority of people with left hemiplegia do not develop a somatoparaphrenic delusion; and mirror agnosia does not necessarily result in a mirrored-self delusion (Coltheart *et al.*, 2011). To buttress a cognitive account of delusion formation, it is necessary to suppose that sufferers have an additional deficit in evaluating the likelihood of their beliefs. This 'two-deficit account' has been promoted especially by Coltheart and colleagues (e.g. Coltheart, 2007; Coltheart *et al.*, 2011). The proposal is that, in addition to what we may call their 'primary' lesion, deluded individuals also have a second impairment of the right hemisphere, possibly involving the lateral prefrontal cortex. If the hypothesized deficit in belief evaluation is provisionally accepted, then an explanation is required for why it results only in a monothematic delusion rather than generating a proliferation of bizarre beliefs. Coltheart *et al.* (2011) suggest that this is so for two reasons. First, the belief evaluation system is not abolished but only impaired. A somatoparaphrenic delusion can be temporarily eliminated by introducing cold water into the left auditory canal of the sufferer, a stimulus known to increase right hemisphere activation, so possibly serving to 'rev up' the impaired belief evaluation system located there. Secondly, this impaired evaluation system is adequate for rejecting most bizarre ideas that might suggest themselves to a person, but it fails when there is continuously or repeatedly present data to support a bizarre idea. A somatoparaphrenic delusion is continually supported by the inability to feel or move the paralysed limbs, and a Capgras delusion is repeatedly reinforced by encounters with the spouse who does not elicit a normal autonomic response.

The extent to which the two-deficit theory of monothematic delusions will find further empirical support remains to be seen. It serves, however, as a good example of what a cognitive neuropsychiatric account of a 'psychotic' symptom can look like.

SUMMARY OF SECTION 4

- Psychosis is not a unitary construct, and has a large overlap with cognitive dysfunction.
- Whilst the literature gives a good indication of where cognitive deficits lie in psychosis (e.g. IQ), it is clear that there are also cases of schizophrenia and psychosis that do not demonstrably display cognitive impairment.
- Experiential distortions in schizophrenia have been described in cognitive terms, for instance 'jumping to conclusions reasoning'; however, notions such as hypersalience and disconnectivity syndrome both hold a degree of usefulness for conceptualizing schizophrenia.
- The clinical manifestations reflect the observation that delusions and other psychotic phenomena arise in part from attempts by the sufferer to make sense of a seemingly disorganized 'hypersalient' world with reduced or disturbed cognitive capacity to address these extra demands.
- Monothematic delusions have proved to be an important 'test-bed' for the notion of a cognitive account of a psychiatric symptom.

5 DEPRESSION

The essential feature of a depressive disorder is a period of at least 2 weeks during which there is either depressed mood *and/or* reduction in capacity for enjoyment (anhedonia) (DSM-IV; American Psychiatric Association, 2000). The individual must also experience at least four other symptoms drawn from a list that includes, amongst others, feelings of guilt, difficulty thinking, concentrating, or making decisions, and reduction or decrease in energy or activity levels (anergia). The symptoms must persist for most of the day nearly every day and be accompanied by clinically significant distress and impairment in social or occupational functioning. It is of note that, by virtue of these two core symptoms, the definition of 'depression' allows for a diagnosis in the absence of alteration of mood. This may be both beneficial and potentially problematic in terms of consideration of the occurrence of the disorder in different organic conditions. On a pragmatic level, many organic conditions labelled as 'depression' appear to respond to treatment (both pharmacological and psychological) regardless of aetiology. However, depression is a heterogeneous condition. The variability of organic pathology in conditions that contribute to depression adds to problems in the study of the depression syndrome.

5.1 Emotion

The loss of a job is a classic example of an event that has the potential to cause considerable variability in emotion. The interpretation of such events is crucial, as they may be 'depressogenic' to one individual but have the completely different effect of inspiring 'happiness' in another. In cognitive terms, this may be explained by one individual taking the view that job loss is bad news and developing subsequent negative thoughts relating to their self-esteem and future. However, another person may see it as an opportunity for a fresh start, and take an optimistic view (you will find a similar discussion of the importance of cognitive appraisal in the generation of emotion in Chapter 16). Important cognitive factors relevant to this process include consideration of how negative affect may be suppressed. Studies have employed fMRI imaging in the context of cognitive 'reframing' of events. Converging evidence points towards ventral-frontal areas of the brain being involved in changes in affect, as well as in the capacity to regulate emotion and cope with distress (Taylor *et al.*, 2003). In addition, voluntary suppression of negative emotion in healthy individuals has been found to recruit limbic and paralimbic brain regions including the amygdala and prefrontal cortex (Phan *et al.*, 2005). Of note is a possible dissociation between anatomical regions involved in 'feeling' versus 'evaluating' emotional events. The hypothalamus, cingula, and insula are thought to 'feel' emotional events and process emotion-laden states. Other structures, usually the amygdala (Damasio, 1999) are thought to 'index' the occurrence of an emotional event but not necessarily contribute to emotional feeling. Patients with amygdala damage do not differ in their phenomenological experience of events but may attach less emotional significance to them (Vuilleumier *et al.*, 2004).

5.2 Neuroanatomical correlates of depression

One of the seminal papers looking at the functional neuroanatomy of the emotional modulation of cognitive processing in depressed patients is described in Box 15.4.

Neuroimaging studies implicate the areas of prefrontal, cingulate, and limbic structures (Dolan *et al.*, 1994; Drevets *et al.*, 1997) in depression, with most consistent findings for ventromedial abnormalities. These areas intercommunicate with structures involved in emotion and autonomic responses to psychological insult, which play a central role in associating emotional experiences with decision making. However, studies investigating the specific brain regions involved have yielded mixed results. Meta-analysis of fMRI studies has revealed that patterns of abnormality are not necessarily consistent and few studies have combined neuropsychological and functional imaging paradigms to investigate the correlation between brain

BOX 15.4 RESEARCH STUDY Elliott *et al.* (2002)

Elliott and colleagues took 10 depressed patients and 11 healthy volunteers. They performed an emotional go/no-go task whilst undergoing functional magnetic resonance imaging (fMRI). In separate blocks of trials, they pressed a button in response to one of happy, sad, or neutral words but did not respond to words in the other two emotional categories. The task allowed comparisons of the neural responses to happy, sad, and neutral words.

In a region of the brain known to be associated with depression (the dorsolateral prefrontal cortex), fMRI findings indicated that depressed patients showed an enhanced neural response to sad words compared to healthy controls. The dorsolateral prefrontal cortex is known to be involved in higher cognitive functions, such as working memory, episodic memory retrieval, attentional set shifting, and planning. The increased response to sad targets suggests a distinct system for the processing of mood and a preference for the processing of material whose content is in keeping with a depressed patient's emotional state. In other words, a depressed person appears to be better able to attend and process sad material. You will find a similar discussion of such cognitive biases in Chapter 16.

areas and prefrontal functioning (Rogers *et al.*, 2004). There may be a number of reasons for these inconsistent results, ranging from methodological variability among the studies themselves to clinical heterogeneity of the depressive groups. Perhaps most likely is the possibility that depression is not the result of abnormality within a single brain area, but the result of multiple system processes affecting discrete but functionally interlinked pathways.

5.3 Depression phenotype and how it relates to cognition

What, then, of depression clinically? How does one relate the concept of a dimensional approach often adopted in neuropsychology with the categorical approach afforded by a diagnosis of major depression in psychiatry?

5.3.1 Mood

In cognitive terms, Elliott *et al.*'s (2002) study, along with others, demonstrates an enhanced neural response to sad stimuli in depressed individuals. The observation of a greater capacity for depressed individuals to 'capture' sad material is consistent with a theory of depression in which depressed people find it increasingly difficult to 'exit' a vicious spiral. Emotional disturbance leads to an attentional bias towards sad events, which leads to increased salience of negative material compared to healthy controls. See also Chapter 16 for a similar discussion of cognitive biases.

5.3.2 Anhedonia

Regarding anhedonia, there is increasing recognition of the role of the mesolimbic dopamine system in depression, given the role of anhedonia or altered reward processing as a central feature of the condition. Dopaminergic mechanisms in the ventral tegmental area and nucleus accumbens are thought to play a role in the mediation of natural rewards, and indeed it has been suggested that the phenomenon of hedonic homeostatic dysregulation syndrome (Giovannoni *et al.*, 2000), a condition whereby Parkinson's individuals self-administer medication in excess of levels prescribed by their physician, in fact reflects patients' attempts to treat their low mood rather than achieve an elevation of mood (Okai *et al.*, 2011).

5.3.3 Anergia

Energy levels are less easily linked to a pathological state. A framework has been suggested that involves an integrated evaluation of expected rewards and energetic costs associated with continued performance (Boksem and Tops 2008), implying an overlap with the frameworks suggested for both low mood (attentional bias causing a tendency to focus on costs) and an undervalued expectation and experience of reward. This results in reduction in motivation.

5.4 Stroke

The prevalence of depression is high following stroke (30 per cent), especially in the first few weeks after the event (Robinson *et al.*, 1984). Additionally, and unsurprisingly, rates of cognitive impairment are high following stroke (Hackett and Anderson, 2005; Hackett *et al.*, 2005).

Studies relating stroke, depression, and cognitive impairment have generated mixed results. Some have shown such an association (Bolla-Wilson *et al.*, 1989; Desmond *et al.*, 2003; Robinson *et al.*, 1986; Spalletta *et al.*, 2002), whilst other studies have failed to do so (Dam, 2001; Madureira *et al.*, 2001). Still others have suggested left hemisphere-damaged patients with major depression are more likely to exhibit cognitive deficits (Narushima *et al.*, 2003; Spalletta *et al.*, 2002). Cognition is usually assessed by a score on the Mini-Mental State Examination (MMSE). However, Bolla-Wilson *et al.* (1989) administered a variety of tests of executive function, language, and temporal orientation, and again demonstrated a cognitive deficit in those with major depression and left lesions, in comparison to those with left lesion alone. Furthermore, social factors such as living alone, place of residence, social support, and social isolation appear consistently associated with depression following stroke, but factors such as prior history of depression have not been shown to be. This lack of an association may be partly accounted for by the low number of studies that record this as a variable in data collection and analysis (Hackett and Anderson, 2005).

The profile described above relates to single vascular events. Evidence of vascular change from MRI and post-mortem investigations has contributed to the idea of a 'vascular depression' (Taylor *et al.*, 2004). This proposes a depression that appears to have some phenomenological differences from 'non-vascular' depression, being characterized by less depressive ideation (i.e. guilt) but more psychomotor retardation, apathy, lack of insight, and functional disability (Alexopoulos *et al.*, 1997). However, repeated ischemic (involving restricted blood supply) events are also known to be relevant to the aetiology of vascular dementia and to have a heavy overlap with risk factors for Alzheimer's dementia in later life (Forette *et al.*, 1998). This overlap adds to the difficulty in attempting to describe a specific cognitive profile relating to repeated vascular events. The traditional divide has been around the assumption that Alzheimer's disease leads to a 'general' decline in global cognitive profile, whilst cognitive testing in vascular disease demonstrates areas of decline as well as areas of relative sparing of cognitive function (i.e. a 'patchy' spread of cognitive deficits).

Sheline and colleagues do attempt to quantify vascular involvement to a degree with a range of neuropsychological tests for those who meet criteria for depression in late life (Sheline *et al.*, 2006). They found deficits in the five cognitive domains they assessed (working memory, language processing, episodic memory, executive function, and processing speed). Using a hierarchical regression model, subject age, depression severity, education, and vascular changes on MRI were significantly correlated with performance in each of the cognitive domains. Processing speed variance, closely followed by executive function, proved significant predictors of cognitive function. Their results appear to indicate that deteriorations in cognitive function are best explained by core deficits in cognitive processing speed, and that executive function may be a mediator of this.

A study by the same group prospectively followed up a large number (n = 217) of patients aged over 60 with major depression (Sheline *et al.* 2010). Using fMRI, they found that an increase in the number and severity of vascular changes (distributed across the cortex and sub-cortex) was associated with depression severity (as assessed by the Montgomery-Asberg Depression Rating Scale) and cognitive deficit (namely processing speed, executive function, and language skills). The amount of vascular change apparent on imaging also appeared to predict antidepressant treatment response. It was hypothesized that the vascular contribution is an effect on frontal white matter integrity, which then results in deficits in neuropsychological functioning and mood regulation, i.e. the same aetiological mechanism may account for both sets of findings.

5.4.1 What about depression in other conditions?

Parkinson's disease may be an ideal condition to study in order to gain a greater understanding of the relationship between neuropathology, depression, and cognition, by virtue of the high prevalence of Parkinson's sufferers in the population. A recent paper by Brown and colleagues aimed to better understand the phenomenology of depression and anxiety in Parkinson's in an attempt to identify whether it is a heterogeneous construct (Brown *et al.*, 2011). Using a form of factor analysis, four subgroups were identified (see Table 15.4).

Group II showed more depressive features than the other groups, with a greater proportion (52.2 per cent) scoring below the cut-off for cognitive impairment on the Addenbrooke's Cognitive Examination (ACE), a validated tool for the assessment of dementia providing scores for differing cognitive domains such as executive function. The depressive phenotype of Group II appeared to have a characteristic profile of what have been termed 'biological' or somatic symptoms. This may indicate a biological correlation with depressive phenotype and cognition. What is striking is the similarity with the phenotype described in the vascular group above (Section 5.4), indicating perhaps an integral involvement of sub-cortical and associated fronto-striatal regions. In contrast, Group III, which had the fewest cognitive deficits, was notable for the absence of

TABLE 15.4 Depression subtypes in Parkinson's disease

Clusters	Common symptoms	Uncommon symptoms
Group I (anxious-depressed) n = 44	As Group II, but also: • Prominent worry • Tension/restlessness • Irritability • General anxiety/panic • Fear • Autonomic symptoms	• Anhedonia
Group II (depressed) n = 46	• Sad/empty mood • Loss of interest • Slow thinking • Loss of confidence • Lack of energy • Poor concentration	• Anhedonia • Worthlessness • Hopelessness • Guilt
Group III (anxious) n = 113	• Prominent worry • Tension/restlessness • Moderate probability of severe anxiety-related symptoms	• Depressive symptoms
Group IV (healthy) n = 310	• No prominent symptoms	• No symptoms

depressive symptoms. There may be two potential reasons for such a finding in Group III. The first is that this group had a higher degree of cortical pathology in addition to the more usual sub-cortical changes seen in Parkinson's disease (the conditions can also affect cortical areas, and some have argued that the amount of cortical involvement defines where along the spectrum from Parkinson's dementia – more sub-cortical – to dementia with Lewy bodies – more cortical – an individual may lie). The second reason may be based on the observation that this group had a higher levodopa load than many of the other groups. Levodopa is prescribed by the neurologist in order to treat the motor component of the illness. Perhaps this group needs a higher dose of dopaminergic stimulation in order to address the cognitive (and depressive) component of their illness.

5.5 Traumatic brain injury

Greater attention has been focused on the relationship between lesion location and depression in stroke than in traumatic brain injury (TBI). This is perhaps related to the difficulty in localizing anatomical regions damaged by TBI.

However, Jorge and colleagues prospectively followed up a group of patients (n = 91) with traumatic brain injury, along with a group (n = 27) who had multiple trauma without evidence of central nervous system involvement (Jorge *et al.*, 2004). All patients were followed up at 3, 6, and 12 months. The two groups had comparable traumatic injuries and functional impairment (using validated scales of injury and functional independence). Significantly more brain injury patients (33 per cent) developed symptoms consistent with major depression during the first year than did non-brain injured patients (7.4 per cent). Depressed individuals also had significantly worse executive functioning as measured by the numbers of preservative errors in the Wisconsin card sorting test and the completion times for the Trail making tests. In addition, those with history of 'mood disorder' were found to be more likely to develop depression.

A further feature of TBI is an indication that depression following traumatic injury may be related to insight or awareness of disability such that those with poor insight have less depression (Amador *et al.*, 1994; David, 1990; Prigatano and Schacter, 1991). Ponsford and colleagues report on a longitudinal sample of closed head injury patients, demonstrating that an improvement in insight coincides with onset of a dysphoric mood state (depression, anxiety, and reduced self-esteem) (Ponsford *et al.*, 1995). These results appear to indicate that as the individual becomes more self-aware their mood worsens. However, there are inconsistencies in how deficits of insight are associated with neuropsychological and injury variables (Bogod *et al.*, 2003; Prigatano and Schacter, 1991).

So what are the links between brain injury, cognitive impairment, and mood? They are often non-specific, such as in the example in Box 15.5, where the brain injury was thought to have occurred because of an obstruction in one of the arteries that supply oxygen to the brain. However, there may be direct or indirect relationships between the conditions. For instance, those with a history of depression may be more likely to sustain brain injury (for example, due to self-injury) and therefore develop cognitive impairment. Additionally, those with cognitive impairment may reflect an association between low mood and brain injury (for example, motor vehicle accident associated with alcohol abuse and co-morbid depression). Finally, there may be confounding variables that contribute to both conditions (for example, alcohol misuse may lead to an increased propensity for cognitive impairment, but alcohol misuse is also often co-morbid with low mood). In such cases the traumatic injury would be incidental. Whilst difficult to prove, research should aim to tease out and verify which are the direct associations and which are confounds.

Furthermore, it remains unclear whether emotional disturbances such as depression and anxiety in their own right impair neuropsychological test performance, or whether brain injury predisposes an individual to emotional disturbance. If so, why do so many patients suffering anxiety disorders without brain injury tend to perform normally on cognitive testing? Does neuropsychological impairment cause emotional disorder, and does emotional distress influence performance on neuropsychological testing?

Various methods can be used to address these questions. One is to examine cognitive deficits before and after antidepressant treatment for depression. For example, Fann *et al.* (2001) used this method in examining cognitive deficits in patients with mild traumatic brain injury (n = 15). Of these, 66 per cent had full remission of their depressive symptoms and were found to have improvements in psychomotor speed and short-term verbal and visual memory. Another method would be to compare individuals with both major

BOX 15.5 CASE STUDY Adapted from Trimble and Cummings (1981)

A 58-year-old male barrister was admitted to hospital after being found unconscious at work. On examination he was lethargic with occasional periods of arousal.

He gradually recovered and was discharged from the hospital. At home, a marked change in his behaviour compared to his pre-morbid activity was noted. He was often unresponsive, lacked motivation, and spent much of his time lying down or sleeping. He tended to overeat, put on weight, and occasionally wet himself. There was a change in mood with lability and tearfulness, and his partner reported that he was often low in mood and this was not amenable to reassurance.

Although during the war the patient had been discharged from the army with a label of ‘nervous breakdown’, he had no further psychiatric problems after leaving the service, and had been steadily employed in his profession.

On examination 6 months after discharge from the hospital, he was emotionally labile and des-cribed lethargy and depressive and persecutory feelings with suicidal ideation. His speech was slow and deliberate, but there was no dysphasia.

MRI demonstrated recent strokes in the subcortical region. Formal neuropsychological evaluation revealed slowness and difficulty in sustaining attention. He had a verbal/performance discrepancy of 38 points on the Wechsler Adult Intelligence Scale (WAIS), and memory for visual and verbal material was poor.

He responded well to an antidepressant but medication for his delusional belief caused him to sleep even more. It was therefore stopped. He was referred to a day centre where he was able to undertake planned daily activities under supervision. Although this is not an obvious case of traumatic brain injury, because the injury was caused by an arterial obstruction not an external force, it does show the non-specific connections between brain injury and behavioural symptoms.

depression and TBI with those having lesser severities of depression and TBI. It would then be possible to look for cognitive differences among the groups. Rapoport *et al.* (2005) conducted such a comparison and found no difference in severity of injury between groups but poorer cognitive functioning on measures of working memory, processing speed, and verbal memory amongst the more depressed group. Subjects with major depression were also found to have more perseverative errors. These findings remained after controlling for past history of depression and age.

5.6 Pseudodementia

Finally, a proportion of elderly individuals with depression present a syndrome of dementia that subsides on resolution of the depressive symptoms. This group are said to be suffering from pseudodementia. In general terms, they are a group with objective low mood, but their subjective complaint is usually of memory impairment (rather than the low mood). This group of patients may show inconsistent performance on cognitive tests, often because they become anxious when asked to perform tests they perceive as difficult. Factors suggestive of pseudodementia are a past history of depression and abrupt onset of illness, and the patient tends to highlight the memory deficit and typically responds with 'I don't know' when asked questions designed to test their memory (the patient with severe Alzheimer's dementia will often quite freely give a wrong answer). The 'dementia' component often manifests itself in deficits in attention, free recall, motor speed, word fluency, and syntactic complexity (Emery and Oxman, 1992). Patients appear to recover seemingly full cognitive capability when the depressive episode subsides, even when it has been untreated. However, a follow-up study by Alexopoulos *et al.* (1993) of elderly individuals with pseudodementia found that over the course of 3 years the pseudodementia group were almost five times more likely to develop an irreversible dementia.

SUMMARY OF SECTION 5

- Neuroimaging studies in patients with primary depression have found structural and metabolic abnormalities in regions of the prefrontal cortex.
- These brain regions appear related to the cognitive biases that patients with depression experience, for instance an enhanced ability to notice sad words.
- Depression following brain injury is most likely a 'multi-brain-system' process.
- The occurrence of depression in stroke, and in particular the observation that it is more likely to occur in the early stages of stroke, may support the idea that (akin to the process of physical recovery of limb weakness following stroke) there may be an initial failure of brain mechanisms to maintain homeostatic emotional control in a period of brain insult.
- A relatively non-specific link with cognition appears to be emerging from the literature, which likely reflects the fact that several distinct but interconnected systems regulate mood, such as affect, cognition, sleep, appetite, psychomotor activity, emotional regulation, and homeostatic drives.
- When these systems are perturbed, whilst the most obvious manifestation may be a 'mood disorder', there is an association with other neuropsychological functions.
- Social influences (for example work stress) may result in failure of the remaining system to maintain homeostatic emotional control in times of increased cognitive demand, with more research needed in relation to directional influences or the cause-and-effect relationship between cognition and affect.
- The aetiologies of mood disorder may be diverse, but the neural dysfunctions underlying the clinical manifestations are felt to be similar. This conceptualization of brain systems involved in mood disorders may prove useful for research and testing.

6 DISINHIBITORY PSYCHOPATHOLOGIES

This section deals with neuropsychological profiles in obsessive–compulsive disorder (OCD), hoarding, and impulse control disorders (ICDs) in those with brain disease. Investigation into phenomenological differences between these conditions and their association with discrete pathologies can potentially add to our understanding of them.

These disorders are included together in this section because they all appear to have at their core a lessened control over response inclinations. Nevertheless, there are some important differences between the disorders. In some, patients appear driven by the pursuit of pleasure or gratification (ICDs) whilst in others they appear driven by an effort to avoid anxiety (e.g. OCD). It would seem that these characteristics (which can be conceptualized as deficits in decision making) may be well fitted to a neuropsychological framework and to our understanding of cognition.

6.1 Obsessive–compulsive disorder

Obsessive–compulsive disorder (OCD) consists of two core features: intrusive distressing thoughts that are recognized as one's own, and repetitive, compulsive acts that usually arise in an attempt to neutralize the thoughts (see DSM-IV, American Psychiatric Association, 2000). It has a prevalence of some 2–3 per cent of the population, yet has been relatively understudied in comparison to many other neuropsychiatric conditions. In this section we aim to bring together what is known of the cognitive and biological components of the disorder.

There is much to lend support to the concept of 'organicity' of this condition. In animal models, repetitive, stereotyped behaviours (e.g. circular pacing) following the administration of dopamine (Graybiel and Rauch, 2000) have been interpreted as 'compulsive' in their behavioural phenotype. Additionally, the development of obsessive–compulsive behaviours has been noted following traumatic brain injury in sufferers with no history of OCD prior to the injury (Stengler-Wenzke *et al.*, 2003). There is an association with Tourette's syndrome, other tic disorders, and possibly Huntington's disease (Cummings and Cunningham, 1992; Mell *et al.*, 2005). Finally, there is evidence of a causal association between OCD symptoms and a bacterial infection resulting in what is known as the 'paediatric autoimmune neuropsychiatric disorder' (PANDAS) (Snider and Swedo, 2004). PANDAS sufferers have been reported to display writhing, and jerky and wriggling movements, which has led to speculation that basal ganglia are implicated in the disorder, since this is the part of the brain that is linked with the initiation and suppression of movements.

OCD has high co-morbidity with many other disorders, such as depression, other anxiety disorders like body dysmorphic disorder, and trichotillomania (a compulsive disorder marked by repeated pulling out of one's hair). Some have proposed that disinhibition may underlie both thought and behaviour in OCD, with the cognitive symptoms (such as intrusive thoughts and mental rituals) and the behavioural symptoms (such as washing and checking rituals) both involving a failure to inhibit or shift attention away from these repetitive phenomena (Tolin *et al.*, 2002).

Neuroimaging findings in OCD are perhaps the most replicated in the psychiatric literature and reveal physical and functional abnormalities in sub-cortical structures. Mataix-Col and colleagues (2004) have contributed a body of work on this condition indicating different symptom domains in OCD that have differing and distinct neural correlates. The putamen, globus pallidus, and thalamus are involved in checking symptoms; the bilateral ventromedial prefrontal cortex and right caudate are implicated in washing symptoms; and the left pre-central gyrus and orbitofrontal cortex play a part in hoarding behaviour (Mataix-Cols *et al.*, 2004). PANDAS individuals also appear to have MRI changes in the basal ganglia. The possible involvement of the basal ganglia, and the observation that impaired inhibitory processes seem to be central to the symptoms and neuropsychology of OCD, suggest that there is clear merit in investigating whether inhibitory failures contribute to any aspects of cognitive deficit in OCD, specifically with consideration as to whether the process is implicated in thoughts or actions, or indeed both.

Veale *et al.* (1996) looked for evidence of slowness of thought and impaired executive functioning in OCD sufferers, akin to similar symptoms demonstrated in sub-cortical organic conditions (see Section 2.5). They used the attentional set shifting task and the Tower of London test (see Chapter 10), with OCD subjects being matched in relation to IQ with healthy controls. No specific memory deficits were detected compared to controls. In the Tower of London test, whilst no differences were found in the total number of errors, differences were found for time spent in correction of such errors. OCD sufferers seemed to have a deficit in generating alternative strategies when they had made a mistake. They also appeared impaired in their ability to learn from errors in the set shifting task. Overall, this may suggest that OCD patients encode important information more poorly than controls and encode too much irrelevant information, leading them to become more distractible and to show increased perseveration with (repetition of) incorrect strategies. In other words, the findings reflect attentional problems in these individuals and/or a failure to appropriately encode information relevant to the self-selection of actions. Savage *et al.* (2000) used the Rey-Osterrieth Complex Figure Test (requiring subjects to draw a complex interlinked figure and then redraw it following a period of delay to assess 'non-verbal' recall; Rey, 1941) to look at non-verbal memory in OCD individuals. Their findings appeared to suggest a failure to encode the most relevant information, but what was encoded stayed in memory and was retrieved. However, patients appeared to use less systematic organizational strategies during encoding and therefore learned less than controls. When external organization was employed (semantically cued recall on a verbal learning task), there was an improvement in this apparent frontostriatal deficit. Savage *et al.* speculated that the failure in encoding relevant information could lead to fragmented memory for repetitive episodes, instances of which are difficult to differentiate. So the OCD individual might think, 'Did I lock the door?', 'Was that today or yesterday?', and 'How can I be sure which?'. Purcell and colleagues obtained similar findings using the Cambridge Neuropsychological Test Automated Battery (Purcell *et al.*, 1998; Robbins *et al.*, 1994), purported to be discriminatory between the cognitive profile of early sub-cortical and cortical disease. In addition, apparent spatial working memory deficits have also been discovered (Van Der Wee *et al.*, 2003). The findings appear to reflect the clinical phenotype of OCD, namely that of doubting one's own memory and mental slowness (due to internal errors and perseveration, which is reminiscent of a sub-cortical profile).

A minority of patients with OCD do not respond to any medical or psychological treatments (known as treatment-resistive OCD). If their distress is of sufficient severity, they may be judged to warrant what is known as psychosurgery. This usually involves removal of a frontal part of the brain. One study compared OCD sufferers following surgery with unoperated-on controls. The psychosurgery group were highly heterogeneous in terms of anatomical region of brain operated on. The study found mixed improvement in OCD symptoms following surgery and preserved neuropsychological functioning on the whole (Hay *et al.*, 1993). Those with involvement of ventral striatum and ventromedial frontal areas did poorly on the Wisconsin card sort task; those with dorsolateral lesions did poorly on Performance IQ.

However, there are several limitations to the work on neuropsychological functioning in the OCD group. Many studies have failed to control for co-morbidity, failed to match IQ in subjects, and failed to take into consideration the varying degrees of duration and severity of the OCD condition. Conclusions were also drawn that at times could more plausibly be applied to the fact that many OCD individuals also suffer from depression. Additionally, there has long been a discussion of the ecological validity of many of the tests used (in particular, those of executive functioning). It is apparent that OCD sufferers, even where there is a clear association with neurological illness, do not appear to suffer hugely from cognitive deficits, whilst at the same time still being very functionally impaired. What is unclear is how much the disability is associated with the illness and how much it is due to secondary factors such as specific, cognitive deficits.

6.2 Hoarding

Collecting and saving possessions is a basic animal instinct that may have some evolutionary function. Like other behaviours classified under the term 'neurosis', hoarding behaviours are part of a spectrum of behaviours that includes 'the norm'; such behaviours only become defined as pathological when they impair social or occupational functioning. Hoarding does not have a formal identification in the DSM-IV and has classically been subsumed within the category of OCD. It is, however, seen in a wide range of psychiatric and neurological conditions such as dementia, OCD, schizophrenia, autism, and following traumatic brain injury. It is also becoming increasingly clear that pathological hoarding can occur in its own right as an independent condition (Mataix-Cols *et al.*, 2011).

Hoarding lends itself well to study in animal models as it is a naturally displayed form of behaviour (e.g. found in birds, rodents, and primates). The ventromedial prefrontal cortex and sub-cortical limbic structures have been implicated. Indeed, hoarding behaviour has been induced in rats by stimulation of the lateral hypothalamus (a region that also promotes control of food intake and sexual behaviour; see Section 6.3 on ICD below). Similarly, the dopaminergic system has been implicated in hoarding behaviour, with disappearance of hoarding behaviour being observed with induced lesions of the mesolimbic dopamine neurons in rats, and restoration of such behaviour to control levels by treatment with dopamine (Kelly and Iversen, 1976).

In humans, studies of those who began hoarding following traumatic or vascular lesions predominantly show the involvement of sub-cortical regions (Anderson *et al.*, 2005; Volle *et al.*, 2002). A report on a number of cases by Anderson *et al.* (2005) found no difference between hoarding (n = 9) and non-hoarding participants (54 brain injury controls) in terms of verbal and non-verbal intelligence, but trends towards significance in the Tower of Hanoi task and memory tasks. The implication may be that there are two main reasons for the clinical manifestations of hoarding: 1) an executive planning deficit that may result in a failure to discard objects; and 2) a memory deficit that may contribute to excessive acquisition of objects. These results will, however, need to be replicated in a larger sample.

Recently, an association between hoarding and dopaminergic medication use has been demonstrated in Parkinson's disease (O'Sullivan *et al.*, 2010), a condition that is often associated with other repetitive behaviours (e.g. pathological gambling, hypersexuality). The authors propose that the dopaminergic medication causes patients to overestimate the intrinsic value of the items in their possession and thus retain and accumulate them.

6.3 Impulse control disorders

Impulse control disorders (ICDs) constitute a group of conditions that have at their core a failure to resist a drive or temptation to perform an act that is harmful (see DSM-IV; American Psychiatric Association, 2000). Examples include pathological gambling and trichotillomania (compulsion to pull one's hair out). There appears to be a disruption of active inhibitory processes regulating rewarding behaviour. In other words, disinhibition is a process where the 'cap' is removed from certain drives and motivations and they have 'free rein'. It has been associated with ventral sub-cortical structures (Cummings, 1995).

In the preparation of the upcoming fifth edition of DSM, there has been consideration of whether to group ICDs under the obsessive–compulsive disorders (OCDs) or in a parallel category of behavioural addictions (Hollander and Allen, 2006). Both categories are manifest by an inability to refrain from repetitive acts when presented with immediate reinforcement, despite long-term negative consequences. The key difference between the two groups is the function of the repetitive behaviours. Behavioural addictions are driven by the pursuit of pleasure, arousal, or gratification, whereas in OCDs they are driven by an effort to avoid anxiety. Marks likened the contrast between the two as a 'push versus pull' effect, with addictions indicating an attraction towards a certain state and compulsions denoting avoidance of discomfort (Marks, 1990).

As mentioned above, the lessened control of response inhibition also impacts on long-term goals. This has been termed a 'myopia (visual short-sightedness) for the future' (Bechara *et al.*, 2000). It is unsurprising that the frontal lobes and associated structures are implicated given their involvement in willed intentional action, inhibition of pre-potent tendencies, reasoning, and self-awareness.

Cloninger (1987) subsumes impulsivity into one of four dimensions that he considers make up the personality trait of 'novelty seeking' (the others being exploratory excitability, disorganization, and extravagance). It is contrasted with another temperamental trait of 'harm avoidance' (Cloninger, 1987). If ICDs are conceptualized as a myopia for the future, harm avoidance sufferers have the opposite problem – a hyperopia (or overly focused viewpoint) of the future, manifest in over-anticipation of long-term problems. This has been associated with serotonergic function, highlighting the reciprocal relationship between dopamine and serotonin. Barrett has identified three factors felt to contribute to impulsivity, namely inattention, motor impulsivity, and a non-planning impulsivity, i.e. failure to consider the future.

6.3.1 ICDs in Parkinson's disease

As a term, ICD has been adopted for use in a range of organic conditions, such as Parkinson's disease, and is felt to be related to the use of dopaminergic medication to treat the movement disorder problems (termed PD-ICDs; see Box 15.6 for an example). It affects 14 per cent of sufferers (Weintraub *et al.*, 2010), and includes pathological gambling, compulsive shopping (learned rewards), hypersexuality, and compulsive eating (natural rewards). An intriguing study by Bódi and colleagues assessed novelty seeking in newly diagnosed PD sufferers before and after dopaminergic treatment (Bódi *et al.*, 2009). Novelty seeking was found to be significantly lower in the pre-treatment PD patients compared to healthy controls. After 12 weeks of dopaminergic treatment it was found to be significantly higher compared to healthy controls and also in comparison to the pre-treatment scores. This is in keeping with reports that suggest there is generally a lower level of risk-taking behaviour in PD patients, for instance a lower propensity to smoke or drink caffeine.

One potential model to account for impulse control disorders may relate to the observation of several differing dopaminergic pathways in the brain. Braak has suggested that neurodegeneration in PD follows several stages, from initial PD stages of rigidity, tremor, and akinesia to later stages of apathy and dementia (Braak *et al.*, 2003). The motor cortex degenerates relatively early, with associated movement disability. However, other dopaminergic pathways (such as reward pathways) are relatively intact at this early stage. The medication aimed at treating the movement disability also acts on the reward pathways, essentially causing a relative 'overdose' of reward activation and increasing the chances of becoming addicted to reward-based behaviours. Blood oxygen level-dependent fMRI (see Chapter 14) suggests a phasic firing of dopamine pathways in relation to reward-based stimuli (Bayer and Glimcher, 2005; Schultz *et al.*, 1997). There is increased firing to unexpectedly large rewards and decreased activity to unexpectedly small rewards. In PD patients exhibiting pathological gambling related to dopaminergic medication, a gambling task paradigm with both win and loss outcomes is known to increase ventral striatal dopamine releases to a greater extent than in PD controls (Steeves *et al.*, 2009).

Voon and colleagues used a computational model aimed at allowing a comparison between actual and expected outcomes to look at prediction error in the PD-ICD group (Voon *et al.*, 2010b) in a study where 14 patients had pathological gambling,14 were PD controls, and 16 were healthy non-PD volunteers. They used a probability paradigm in a learning task and found dopaminergic medication to be associated

BOX 15.6 CASE STUDY Mr X

Mr X was a 56-year-old with a 4-year history of Parkinson's disease. He had been retired due to his Parkinson's for 2 years. He was referred to the psychiatrist after reading in the newspaper of a potential link between his medication and gambling. He mentioned he had always 'liked a bet' but since his diagnosis and treatment for his Parkinson's disease it had got increasingly out of hand. He was now £12 000 in debt and his wife had recently moved out of the family home saying she felt the trust had gone from their relationship; he had not told her about the debt until they were threatened with eviction from the family home.

Mr X's score on the MMSE was 30/30, and formal neuropsychometry (an out-patient test of cognition administered by a neuropsychiatrist or neuropsychologist) was within the normal range, with impaired performance shown only on the trails B test (a kind of 'join the dots' task where the dots are presented consecutively). Additionally, there was no evidence of other behaviours known often to be associated with this compulsive gambling (such as compulsive shopping, involving a learned reward, or hypersexuality or compulsive eating, involving natural rewards).

Treatment consisted of a change in his Parkinson's medication to medication that was thought to be less likely to cause this condition. He was referred for psychological work, which involved motivating him to pursue activities he used to enjoy, and introducing strategies to help manage the gambling urges. Additionally, he and his wife met with the neuropsychiatrist, who gave them a talk on the fact that this condition seems to occur in approximately 14 per cent of Parkinson's patients. This led to his wife reconsidering the issues of trust ('It wasn't him, it was the medication, I can see that now.'). However, she continued to complain of him being poorly motivated at home and not being properly responsive to her needs. She appeared to take some comfort in an explanation of her husband's executive difficulties.

with faster learning from gain outcomes along with greater ventral striatal dopaminergic release. This was not seen in the PD controls and the effect was lessened when the patients were off their medication. Additionally, the PD patients with pathological gambling had fewer fMRI changes in relation to loss, indicating less of an ability to learn from negative feedback.

The clinical implications are clear in relation to these findings. Individuals with PD-ICD have a greater than normal tendency to repeat occasionally rewarded behaviours (variable reinforcement) and a reduced tendency towards extinction of these behaviours when they are unrewarded. The study indicates that PD-ICD patients have at their core a dissociation of decision-making processes. It remains unclear what cognitive processes contribute to a vulnerability to these behaviours, and it would be of interest to identify what factors confer such a cognitive dysfunction.

SUMMARY OF SECTION 6

- There is an emerging literature suggesting that OCD may be associated with neuropsychological deficits in comparison to healthy controls.
- There is evidence to suggest only an indirect association with memory deficit and more of an association with cognitive and behavioural inhibitory processes, as well as attentional processes.
- It is possible that the apparent memory deficits could be attributed to association with poor organizational strategies, which may be a reflection of the patient's clinical presentation – they often present with rigid and inflexible rules for life and may be less willing to adopt new strategies to deal with novel problems.

7 CONVERSION DISORDER

Conversion disorder is defined as a condition where neurological signs and symptoms, such as abnormal movements, seizures, or sensory symptoms, are unrelated to an underlying neurological or medical disorder. They are the neuropsychiatrist's 'bread and butter', taking up a large proportion of their workload. In neurology clinics, symptoms that do not fit with an explainable neurological pathology (in other words 'unexplained' neurological illness) account for 30 per cent of cases seen. They are often a source of distress to the patients and are associated with prominent disability. It is important to note that on the whole the sufferer is not feigning these symptoms, but generally believes that their motor function is impaired. An example is described in Box 15.7. Indeed, it is important to note that if the patient is shown to be deliberately feigning their symptoms (e.g. witnessed walking around the room when they believe themselves to be unobserved), then they do not fulfil the criteria for conversion disorder. One hypothesis about the causes of conversion disorder suggests that the loss of function, or paralysis, results from an impairment in the 'generation' of intentions to move, or from impairment in the 'execution' of movement (Spence *et al.*, 2000). The individual says and thinks, 'I cannot'; to an observer it may look like 'I will not', but the reality may be 'I cannot will'.

Whilst there is no specific evidence for cognitive deficits in such a group, there are certain features that are of relevance. Firstly, patients with conversion disorder symptoms in remission have demonstrated failure to habituate to unconditioned stimuli (Horvath *et al.*, 1980). The authors of this study suggest that the failure to habituate may be a risk factor for development of the condition, or that the patients had higher arousal levels in the first instance. Additionally, conversion patients have been shown to have a greater vigilance or attentional bias towards threat stimuli in an emotional Stroop test. Preliminary data from neuroimaging studies implicate a potential link with limbic, amygdala, and motor regions in patients with conversion paralysis. The amygdala is critical for fear conditioning, a form of classical conditioning in which an initially neutral stimulus is presented together with a fear-inducing unconditioned stimulus (e.g. an electric shock). As a result, the neutral stimulus comes to elicit the fear response. Patients with amygdala 'damage' do not acquire conditioned fear responses despite retaining explicit knowledge of the relationship between the unconditioned stimulus and the conditioned stimulus. Voon and colleagues compared healthy volunteers and conversion disorder patients while controlling for depressive and anxiety symptoms. Using fMRI, they found that whereas healthy volunteers had greater right amygdala activity in response to fearful versus neutral stimuli (compared with happy versus neutral), there was no difference in amygdala activity in patients with conversion disorder, suggesting impaired amygdala habituation (Voon *et al.*, 2010a). When compared with the volunteers, conversion patients also showed greater functional connection between the amygdala and motor cortex during fearful versus neutral stimuli.

BOX 15.7 CASE STUDY Mr H

Mr H suddenly developed leg weakness and became unable to walk. The onset was a few months after the death of his mother, who had been bed-bound in the last few months of her life. He is now looked after by his wife, but strangely seemed relatively unconcerned about this sudden-onset disability. He believed the cause to be a stroke and wondered if he had inherited a predisposition to stroke from his mother.

SUMMARY OF SECTION 7

There may be a link between the conversion symptoms and arousal that potentially plays a role in modulating motor networks, so resulting in abnormal motor symptoms.

8 SUMMARY

Despite advances in psychiatric research methods, traditional psychiatric frameworks have offered little towards psychological mechanisms underlying neuropsychiatric conditions such as delusions and hallucinations. Cognitive neuropsychiatry attempts to address this deficit by mapping out the relationship between psychiatric disorders and human cognitive functioning. It also attempts to understand and characterize the cognitive and psychiatric functions of brain structures and their pathology.

Well-described and reliable psychiatric symptoms can increasingly be interpreted in terms of disturbance to normal information-processing systems such as attention, perception, learning, and memory, and be linked to separate but interconnected neural systems. Theoretical accounts of the structure of normal cognitive processes can, in turn, be refined and revised based on empirical findings from patients with different types of psychopathology, in a way similar to the underlying aims of cognitive neuropsychology (see Chapter 13).

Despite the potential advantages of a symptom-based approach in attempting to understand psychopathology, research must avoid making inappropriate assumptions of normality or universality in this field. That is, care is needed in assuming that it is clear what a patient 'would have been expected to do' were it not for their brain injury. Similarly, it cannot straightforwardly be assumed that the patient's cognitive apparatus is representative of all other cognitive systems, or that their deficits reveal a stable underlying process and not a unique and haphazard reorganization of damaged brain systems (Newcombe and Marshall, 1996).

FURTHER READING

David, A., Fleminger, S., Kopelman, M., Lovestone, S., and Mellers, J. (2009) *Lishman's Organic Psychiatry: A Textbook of Neuropsychiatry*, Oxford, Wiley-Blackwell. This text provides a comprehensive overview of neuropsychiatry, containing all of the essential material for those interested in the subject.

Sims, A.C.P. (1995) *Symptoms in the Mind: An Introduction to Descriptive Psychopathology*, London, Saunders Ltd. This text offers a concise and comprehensive introduction to descriptive psychopathology, providing in-depth knowledge of the signs and symptoms of behavioural disorders.

Hodges, J.R. (2007) *Cognitive Assessment for Clinicians*, Oxford, Oxford University Press. This book incorporates advances in recent theory into clinical practice and provides a practical approach to the cognitive evaluation of patients.

REFERENCES

Ala, T.A., Yang, K.H., Sung, J.H., and Frey, W.H. (1997) 'Hallucinations and signs of parkinsonism help distinguish patients with dementia and cortical Lewy bodies from patients with Alzheimer's disease at presentation: a clinicopathological study', *Journal of Neurology, Neurosurgery & Psychiatry*, vol.62, no.1, p.16.

Alexander, G.E. and Crutcher, M.D. (1990) 'Functional architecture of basal ganglia circuits: neural substrates of parallel processing', *Trends in Neurosciences*, vol.13, no.7, pp. 266–271.

Alexopoulos, G.S., Barnett, S., Meyers, R.C., Young, S., Campbell, D., Silbersweig, and Charlson, M. (1997) 'Vascular depression hypothesis', *Archives of General Psychiatry*, vol.54, no.10, pp.915–22.

Alexopoulos, G.S., Meyers, B.S., Young, R.C., Mattis, S., and Kakuma, T. (1993) 'The course of geriatric depression with "reversible dementia": a controlled study', *American Journal of Psychiatry*, vol.150, no.11, p.1693.

Amador, X.F., Flaum, M., Andreasen, N.C., Strauss, D.H., Yale, S.A., Clark, S.C., and Gorman, J.M. (1994) 'Awareness of illness in schizophrenia and schizoaffective and mood disorders', *Archives of General Psychiatry*, vol.51, no.10, p.826.

American Psychiatric Association (2000) *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*, Washington, DC, American Psychiatric Publishing, Inc.

Anderson, S.W., Damasio, H., and Damasio, A.R. (2005) 'A neural basis for collecting behaviour in humans', *Brain*, vol.128, no.1, p.201.

Ballard, C., Ziabreva, I., Perry, R., Larsen, J.P., O'Brien, J., McKeith, I., *et al.* (2006) 'Differences in neuropathologic characteristics across the Lewy body dementia spectrum', *Neurology*, vol.67, no.11, pp.1931–34.

Bayer, H.M. and Glimcher, P.W. (2005) 'Midbrain dopamine neurons encode a quantitative reward prediction error signal', *Neuron*, vol.47, no.1, pp.129–41.

Bechara, A., Damasio, H., and Damasio, A.R. (2000) 'Emotion, decision making and the orbitofrontal cortex', *Cerebral Cortex*, vol.10, no.3, pp.295–307.

Blumer, D. and Benson, D.F. (1975) 'Personality changes with frontal and temporal lobe lesions', *Psychiatric Aspects of Neurologic Disease*, vol.1, pp.151–70.

Bódi, N., Kéri, S., Nagy, H., Moustafa, A., Myers, C.E., Daw, N., *et al.* (2009) 'Reward-learning and the novelty-seeking personality: a between- and within-subjects study of the effects of dopamine agonists on young Parkinson's patients', *Brain*, vol.132, no.9, p.2385.

Bogod, N.M., Mateer, C.A., and Macdonald, S.W.S. (2003) 'Self-awareness after traumatic brain injury: a comparison of measures and their relationship to executive functions', *Journal of the International Neuropsychological Society*, vol.9, no.3, pp.450–8.

Boksem, M.A.S. and Tops, M. (2008) 'Mental fatigue: costs and benefits', *Brain Research Reviews*, vol.59, no.1, pp.125–39.

Bolla-Wilson, K., Robinson, R.G., Starkstein, S.E., Boston, J., and Price, T.R. (1989) 'Lateralization of dementia of depression in stroke patients', *American Journal of Psychiatry*, vol.146, no.5, p.627.

Braak, H. and Braak, E. (1997) 'Frequency of stages of Alzheimer-related lesions in different age categories', *Neurobiology of Aging*, vol.18, no.4, pp.351–7.

Braak, H., Del Tredici, K., Rüb, U., de Vos, R.A.I., Jansen Steur, E.N.H., and Braak, E. (2003) 'Staging of brain pathology related to sporadic Parkinson's disease', *Neurobiology of Aging*, vol.24, no.2, pp.197–211.

Brighetti, G., Bonifacci, P., Borlimi, R., and Ottaviani, C. (2007) 'Far from the heart, far from the eye: evidence from the Capgras delusion', *Cognitive Neuropsychiatry*, vol.12, pp. 189–97.

Broca, P. (1861) 'Perte de la parole, ramollissement chronique et destruction partielle du lobe antérieur gauche du cerveau', *Bulletin de la Société Anthropologique*, vol.2, pp.235–8.

Brown, R.G., Landau, S., Hindle, J.V., Playfer, J., Samuel, M., Wilson, K.C., *et al.* (2011) 'Depression and anxiety related subtypes in Parkinson's disease', *Journal of Neurology, Neurosurgery & Psychiatry*, vol.82, no.7, p.803.

Brugger, P., Regard, M., and Landis, T. (1996) 'Unilaterally felt presences: the neuropsychiatry of one's invisible doppelganger', *Cognitive and Behavioral Neurology*, vol.9, no.2, p.114.

Bruscoli, M. and Lovestone, S. (2004) 'Is MCI really just early dementia? A systematic review of conversion studies', *International Psychogeriatrics*, vol.16, no.2, p.129–40.

Burns, A., Rossor, M., Hecker, J., Gauthier, S., Petit, H., Möller, H.J., *et al.* (2000) 'The effects of donepezil in Alzheimer's disease – results from a multinational trial', *Dementia and Geriatric Cognitive Disorders*, vol.10, no.3, pp.237–44.

Cervantes, A.N., Rabins, P.V., and Slavney, P.R. (2006) 'Onset of schizophrenia at age 100', *Psychosomatics*, vol.47, no.4, pp.356–9.

Cloninger, C.R. (1987) 'Neurogenetic adaptive mechanisms in alcoholism', *Science*, vol.236, no.4800, pp.410–16.

Coltheart, M. (2007) 'The 33rd Bartlett Lecture: Cognitive neuropsychiatry and delusional belief', *Quarterly Journal of Experimental Psychology*, vol.60, pp.1041–62.

Coltheart, M., Langdon, R., and McKay, R. (2011) 'Delusional belief', *Annual Review of Psychology*, vol.62, pp.271–98.

Cummings, J.L. (1995) 'Anatomic and behavioral aspects of frontal subcortical circuits', *Annals of the New York Academy of Sciences*, vol.769, no.1, pp.1–14.

Cummings, J.L. and Cunningham, K. (1992) 'Obsessive-compulsive disorder in Huntington's disease: 1', *Biological Psychiatry*, vol.31, no.3, pp.263–70.

Dam, H. (2001) 'Depression in stroke patients 7 years following stroke', *Acta Psychiatrica Scandinavica*, vol.103, no.4, pp.287–93.

Damasio, A.R. (ed.) (1999) *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, New York, Heinemann.

David, A.S. (1990) 'Insight and psychosis', *The British Journal of Psychiatry*, vol.156, no.6, pp.798–808.

David, A.S., Malmberg, A., Brandt, L., Allebeck, P., and Lewis, G. (1997) 'IQ and risk for schizophrenia: a population-based cohort study', *Psychological Medicine*, vol.27, no.6, pp.1311–23.

Desmond, D.W., Remien, R.H., Moroney, J.T., Stern, Y., Sano, M., and Williams, J.B.W. (2003) 'Ischemic stroke and depression', *Journal of the International Neuropsychological Society*, vol.9, no.3, pp.429–39.

Dolan, R.J., Bench, C.J., Brown, R.G., Scott, L.C., and Frackowiak, R.S.J. (1994) 'Neuropsychological dysfunction in depression: the relationship to regional cerebral blood flow', *Psychological Medicine*, vol.24, no.4, pp.849–57.

Dolan, R.J., Bench, C.J., Liddle, P.F., Friston, K.J., Frith, C.D., Grasby, P.M., and Frackowiak, R.S. (1993) 'Dorsolateral prefrontal cortex dysfunction in the major psychoses; symptom or disease specificity?' *Journal of Neurology, Neurosurgery & Psychiatry*, vol.56, no.12, pp.1290–4.

Drevets, W.C., Price, J.L., Simpson, J.R., Todd, R.D., Reich, T., Vannier, M., and Raichle, M.E. (1997) 'Subgenual prefrontal cortex abnormalities in mood disorders', *Nature*, vol.386, no.6627, pp.824–7.

Egan, M.F., Goldberg, T.E., Kolachana, B.S., Callicott, J.H., Mazzanti, C.M., Straub, R.E., *et al.* (2001) 'Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia', *Proceedings of the National Academy of Sciences*, vol.98, no.12, p.6917.

Elliott, R. (2003) 'Executive functions and their disorders', *British Medical Bulletin*, vol.65, no.1, p.49.

Elliott, R., Rubinsztein, J.S., Sahakian, B.J., and Dolan, R.J. (2002) 'The neural basis of mood-congruent processing biases in depression', *Archives of General Psychiatry*, vol.59, no.7, pp.597–604.

Ellis, H.D. and Young, A.W. (1990) 'Accounting for delusional misidentifications', *British Journal of Psychiatry*, vol.157, pp.239–48.

Ellis, H.D., Young, A.W., Quayle, A.H., and De Pauw, K.W. (1997) 'Reduced autonomic responses to faces in Capgras delusion', *Proceedings of the Royal Society of London, Series B: Biological Sciences*, vol.264, no.1384, p.1085.

Emery, V.O. and Oxman, T.E. (1992) 'Update on the dementia spectrum of depression', *American Journal of Psychiatry*, vol.149, no.3, p.305.

Fann, J.R., Uomoto, J.M., and Katon, W.J. (2001) 'Cognitive improvement with treatment of depression following mild traumatic brain injury', *Psychosomatics*, vol.42, no.1, pp.48–54.

ffytche, D.H., Howard, R.J., Brammer, M.J., David, A., Woodruff, P., and Williams, S. (1998) 'The anatomy of conscious vision: an fMRI study of visual hallucinations', *Nature Neuroscience*, vol.1, no.8, p.738.

Forette, F., Seux, M.L., Staessen, J.A., Thijs, L., Birkenhäger, W.H., Babarskiene, M.R., *et al.* (1998) 'Prevention of dementia in randomised double-blind placebo-controlled Systolic Hypertension in Europe (Syst-Eur) trial', *The Lancet*, vol.352, no.9137, pp.1347–51.

Fotopoulou, A., Solms, M., and Turnbull, O. (2004) 'Wishful reality distortions in confabulation: a case report 1', *Neuropsychologia*, vol.42, no.6, pp.727–44.

Frith, C.D. (1992) *The Cognitive Neuropsychology of Schizophrenia*, Hove, Psychology Press.

Giovannoni, G., O'Sullivan, J.D., Turner, K., Manson, A.J., and Lees, A.J.L. (2000) 'Hedonistic homeostatic dysregulation in patients with Parkinson's disease on dopamine replacement therapies', *Journal of Neurology, Neurosurgery, and Psychiatry*, vol.68, no.4, pp.423–8.

Goldberg, T.E., Egan, M.F., Gscheidle, T., Coppola, R., Weickert, T., Kolachana, B.S., *et al.* (2003) 'Executive subprocesses in working memory: relationship to catechol-O-methyltransferase Val158Met genotype and schizophrenia', *Archives of General Psychiatry*, vol.60, no.9, p.889.

Grace, J., Malloy, P., and Psychological Assessment Resources (2001) *FrSBe, Frontal Systems Behavior Scale: Professional Manual*, Lutz, FL, Psychological Assessment Resources.

Graybiel, A.M. and Rauch, S.L. (2000) 'Toward a neurobiology of obsessive-compulsive disorder', *Neuron*, vol.28, no.2, p.343–7.

Gunnell, D., Harrison, G., Rasmussen, F., Fouskakis, D., and Tynelius, P.E.R. (2002) 'Associations between premorbid intellectual performance, early-life exposures and early-onset schizophrenia', *The British Journal of Psychiatry*, vol.181, no.4, pp.298–305.

Gussekloo, J., Westendorp, R.G.J., Remarque, E.J., Lagaay, A.M., Heeren, T. J., and Knook, D.L. (1997) 'Impact of mild cognitive impairment on survival in very elderly people: cohort study', *BMJ*, vol.315, no.7115, pp.1053–4.

Hackett, M.L. and Anderson, C.S. (2005) 'Predictors of depression after stroke', *Stroke*, vol.36, no.10, pp.2296–301.

Hackett, M.L., Yapa, C., Parag, V., and Anderson, C.S. (2005) 'Frequency of depression after stroke', *Stroke*, vol.36, no.6, pp.1330–40.

Harrington, C. (2007) 'Onset of schizophrenia at age 100 – or psychosis as the presenting symptom of a cognitive disorder?' *Psychosomatics*, vol.48, no.4, pp.360–1.

Hay, P., Sachdev, P., Cumming, S., Smith, J.S., Lee, T., Kitchener, P., and Matheson, J. (1993) 'Treatment of obsessive compulsive disorder by psychosurgery', *Acta Psychiatrica Scandinavica*, vol.87, no.3, pp.197–207.

Hebb, D.O. (1949) *The Organisation of Behaviour*, New York, Wiley.

Hirstein, W.S. and Ramachandran, V.S. (1997) 'Capgras syndrome: a novel probe for understanding the neural representation of the identity and familiarity of persons', *Proceedings of the Royal Society of London B*, vol.264, pp.437–44.

Hodges, J.R. (2001) 'Frontotemporal dementia (Pick's disease): clinical features and assessment', *Neurology*, vol.56, suppl.4, pp.S6–S10.

Hollander, E. and Allen, A. (2006) 'Is compulsive buying a real disorder, and is it really compulsive?' *The American Journal of Psychiatry*, vol.163, no.10, pp.1670–2.

Horvath, T., Friedman, J., and Meares, R. (1980) 'Attention in hysteria: a study of Janet's hypothesis by means of habituation and arousal measures', *American Journal of Psychiatry*, vol.137, no.2, p.217.

James, W. (1950) *Principles of psychology, Vol. 2*, New York, Henry Holt & Co. (original work published 1890).

Jorge, R.E., Robinson, R.G., Moser, D., Tateno, A., Crespo-Facorro, B., and Arndt, S. (2004) 'Major depression following traumatic brain injury', *Archives of General Psychiatry*, vol.61, no.1, p.42.

Kapur, S. (2003) 'Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia', *American Journal of Psychiatry*, vol.160, no.1, pp.13–23.

Kelly, P.H. and Iversen, S.D. (1976) 'Selective 60HDA-induced destruction of mesolimbic dopamine neurons: abolition of psychostimulant-induced locomotor activity in rats', *European Journal of Pharmacology*, vol.40, no.1, pp.45–56.

Kopelman, M.D. (2002) 'Disorders of memory', *Brain*, vol.125, no.10, p.2152.

Kraepelin, E. (1919) *Dementia Praecox and Paraphrenia*, Edinburgh, Livingstone.

Liddle, P.F. and Barnes, T. R. (1990) 'Syndromes of chronic schizophrenia', *The British Journal of Psychiatry*, vol.157, no.4, pp.558–61.

Liddle, P.F. and Morris, D. L. (1991) 'Schizophrenic syndromes and frontal lobe performance', *The British Journal of Psychiatry*, vol.158, no.3, pp.340–5.

Madureira, S., Guerreiro, M., and Ferro, J.M. (2001) 'Dementia and cognitive impairment three months after stroke', *European Journal of Neurology*, vol.8, no.6, pp.621–7.

Marks, I. (1990) 'Behavioural (non-chemical) addictions', *British Journal of Addiction*, vol.85, no.11, pp.1389–94.

Mataix-Cols, D., Pertusa, A., and Snowdon, J. (2011) 'Neuropsychological and neural correlates of hoarding: a practice friendly review', *Journal of Clinical Psychology*.

Mataix-Cols, D., Wooderson, S., Lawrence, N., Brammer, M.J., Speckens, A., and Phillips, M.L. (2004) 'Distinct neural correlates of washing, checking, and hoarding symptom dimensions in obsessive-compulsive disorder', *Archives of General Psychiatry*, vol.61, no.6, p.564.

McKeith, I.G., Ballard, C.G., Perry, R.H., Ince, P.G., O'Brien, J.T., Neill, D., *et al.* (2000) 'Prospective validation of consensus criteria for the diagnosis of dementia with Lewy bodies', *Neurology*, vol.54, no.5, p.1050.

Mell, L.K., Davis, R.L., and Owens, D. (2005) 'Association between streptococcal infection and obsessive-compulsive disorder, Tourette's syndrome, and tic disorder', *Pediatrics*, vol.116, no.1, p.56.

Moritz, S. and Woodward, T.S. (2007) 'Metacognitive training in schizophrenia: from basic research to knowledge translation and intervention', *Current Opinion in Psychiatry*, vol.20, no.6, p.619–25.

Müller, M.J., Greverus, D., Dellani, P.R., Weibrich, C., Wille, P.R., Scheurich, A., *et al.* (2005) 'Functional implications of hippocampal volume and diffusivity in mild cognitive impairment', *Neuroimage*, vol.28, no.4, pp.1033–42.

Narushima, K., Kosier, J.T., and Robinson, R.G. (2003) 'A reappraisal of poststroke depression, intra- and inter-hemispheric lesion location using meta-analysis', *Journal of Neuropsychiatry and Clinical Neurosciences*, vol.15, no.4, p.422.

Newcombe, F. and Marshall, J.C. (1996) 'Idealisation meets psychometrics: the case for the right groups and the right individuals', *Human Cognitive Neuropsychology: A Textbook with Readings*, 351.

Okai, D., Samuel, M., Askey-Jones, S., David, A.S., and Brown, R.G. (2011) 'Impulse control disorders and dopamine dysregulation in Parkinson's disease: a broader conceptual framework', *European Journal of Neurology*, vol.1–5.

Ortiz-Gil, J., Pomarol-Clotet, E., Salvador, R., Canales-Rodreguez, E.J., Sarra, S., Gomar, J.J., *et al.* (2011) 'Neural correlates of cognitive impairment in schizophrenia', *The British Journal of Psychiatry*.

O'Sullivan, S.S., Djamshidian, A., Evans, A.H., Loane, C.M., Lees, A.J., and Lawrence, A.D. (2010) 'Excessive hoarding in Parkinson's disease', *Movement Disorders*, vol.25, no.8, pp.1026–33.

Palmer, B., Dawes, S., and Heaton, R. (2009) 'What do we know about neuropsychological aspects of schizophrenia?' *Neuropsychology Review*, vol.19, no.3, pp.365–84.

Patton, J.H., Stanford, M.S., and Barratt, E.S. (1995) 'Factor structure of the Barratt impulsiveness scale', *Journal of Clinical Psychology*, vol.51, no.6, pp.768–74.

Petersen, R.C., Doody, R., Kurz, A., Mohs, R.C., Morris, J.C., Rabins, P.V., *et al.* (2001) 'Current concepts in mild cognitive impairment', *Archives of Neurology*, vol.58, no.12, pp.1985–92.

Phan, K.L., Fitzgerald, D.A., Nathan, P.J., Moore, G.J., Uhde, T.W., and Tancer, M.E. (2005) 'Neural substrates for voluntary suppression of negative affect: a functional magnetic resonance imaging study', *Biological Psychiatry*, vol.57, no.3, pp.210–19.

Piefke, M., Weiss, P.H., Zilles, K., Markowitsch, H.J., and Fink, G.R. (2003) 'Differential remoteness and emotional tone modulate the neural correlates of autobiographical memory', *Brain*, vol.126, no.3, pp.650–68.

Pomarol-Clotet, E., Oh, T.M.S.S., Laws, K.R., and McKenna, P.J. (2008) 'Semantic priming in schizophrenia: systematic review and meta-analysis', *The British Journal of Psychiatry*, vol.192, no.2, pp.92–7.

Ponsford, J., Sloan, S., and Snow, P. (1995) *Traumatic Brain Injury: Rehabilitation for Everyday Adaptive Living*, Hove, Psychology Press.

Prigatano, G.P. and Schacter, D. L. (1991) *Awareness of Deficit After Brain Injury: Clinical and Theoretical Issues*, New York, Oxford University Press.

Purcell, R., Maruff, P., Kyrios, M., and Pantelis, C. (1998) 'Cognitive deficits in obsessive-compulsive disorder on tests of frontal-striatal function', *Biological Psychiatry*, vol.43, no.5, pp.348–57.

Rapoport, M.J., McCullagh, S., Shammi, P., and Feinstein, A. (2005) 'Cognitive impairment associated with major depression following mild and moderate traumatic brain injury', *Journal of Neuropsychiatry and Clinical Neurosciences*, vol.17, no.1, p.61.

Reichenberg, A., Weiser, M., Rabinowitz, J., Caspi, A., Schmeidler, J., Mark, M., *et al.* (2002) 'A population-based cohort study of premorbid intellectual, language, and behavioral functioning in patients with schizophrenia, schizoaffective disorder, and nonpsychotic bipolar disorder', *American Journal of Psychiatry*, vol.159, no.12, pp.2027–35.

Rey, A. (1941) 'L'examen psychologique dans les cas d'encéphalopathie traumatique. (Les problems.)', *Archives de Psychologie*.

Robbins, T.W., James, M., Owen, A.M., Sahakian, B.J., McInnes, L., and Rabbitt, P. (1994) 'Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers', *Dementia and Geriatric Cognitive Disorders*, vol.5, no.5, pp.266–81.

Robinson, R.G., Bolla-Wilson, K., Kaplan, E., Lipsey, J.R., and Price, T.R. (1986) 'Depression influences intellectual impairment in stroke patients', *The British Journal of Psychiatry*, vol.148, no.5, pp.541–7.

Robinson, R.G., Starr, L.B., and Price, T.R. (1984) 'A two year longitudinal study of mood disorders following stroke. Prevalence and duration at six months follow-up', *The British Journal of Psychiatry*, vol.144, no.3, pp.256–62.

Rogers, M.A., Kasai, K., Koji, M., Fukuda, R., Iwanami, A., Nakagome, K., *et al.* (2004) 'Executive and prefrontal dysfunction in unipolar depression: a review of neuropsychological and imaging evidence', *Neuroscience Research*, vol.50, no.1, pp.1–11.

Savage, C.R., Deckersbach, T., Wilhelm, S., Rauch, S.L., Baer, L., Reid, T., and Jenike, M.A. (2000) 'Strategic processing and episodic memory impairment in obsessive compulsive disorder', *Neuropsychology*, vol.14, no.1, p.141.

Schmahmann, J.D. and Sherman, J.C. (1998) 'The cerebellar cognitive affective syndrome', *Brain*, vol.121, no.4, p.561.

Schultz, W., Dayan, P., and Read Montague, P. (1997) 'A neural substrate of prediction and reward', *Science*, vol.275, no.5306, pp.1593–9.

Shallice, T.I.M. and Burgess, P.W. (1991) 'Deficits in strategy application following frontal lobe damage in man', *Brain*, vol.114, no.2, pp.727–41.

Sheline, Y.I., Barch, D.M., Garcia, K., Gersing, K., Pieper, C., Welsh-Bohmer, K., *et al.* (2006) 'Cognitive function in late life depression: relationships to depression severity, cerebrovascular risk factors and processing speed', *Biological Psychiatry*, vol.60, no.1, pp.58–65.

Sheline, Y.I., Pieper, C.F., Barch, D.M., Welsh-Boehmer, K., McKinstry, R.C., MacFall, J.R., *et al.* (2010) 'Support for the vascular depression hypothesis in late-life depression: results of a 2-site, prospective, antidepressant treatment trial', *Archives of General Psychiatry*, vol.67, no.3, pp.277–85.

Shenton, M.E., Dickey, C.C., Frumin, M., and McCarley, R.W. (2001) 'A review of MRI findings in schizophrenia', *Schizophrenia Research*, vol.49, no.1–2, pp.1–52.

Snider, L.A. and Swedo, S.E. (2004) 'PANDAS: current status and directions for research', *Molecular Psychiatry*, vol.9, no.10, pp.900–07.

Snowdon, D.A. (2003) 'Healthy aging and dementia: findings from the Nun Study', *Annals of Internal Medicine*, vol.139, no.5, part 2, p.450.

Snowdon, D.A., Kemper, S.J., Mortimer, J.A., Greiner, L.H., Wekstein, D.R., and Markesbery, W.R. (1996) 'Linguistic ability in early life and cognitive function and Alzheimer's disease in late life', *JAMA: The Journal of the American Medical Association*, vol.275, no.7, p.528.

Spalletta, G., Guida, G., De Angelis, D., and Caltagirone, C. (2002) 'Predictors of cognitive level and depression severity are different in patients with left and right hemispheric stroke within the first year of illness', *Journal of Neurology*, vol.249, no.11, pp.1541–51.

Spence, S.A., Crimlisk, H.L., Cope, H., Ron, M.A., and Grasby, P.M. (2000) 'Discrete neurophysiological correlates in prefrontal cortex during hysterical and feigned disorder of movement', *The Lancet*, vol.355, no.9211, pp.1243–4.

Sprong, M., Schothorst, P., Vos, E., Hox, J., and Van Engeland, H. (2007) 'Theory of mind in schizophrenia: meta-analysis', *The British Journal of Psychiatry*, vol.191, no.1, p.5.

Staekenborg, S.S., Su, T., van Straaten, E.C.W., Lane, R., Scheltens, P., Barkhof, F., and van der Flier, W.M. (2009) 'Behavioural and psychological symptoms in vascular dementia; differences between small- and large-vessel disease', *Journal of Neurology, Neurosurgery & Psychiatry*, vol.81, no.5, pp.547–51.

Steeves, T.D.L., Miyasaki, J., Zurowski, M., Lang, A.E., Pellecchia, G., Van Eimeren, T., *et al.* (2009) 'Increased striatal dopamine release in Parkinsonian patients with pathological gambling: a [11C] raclopride PET study', *Brain*, vol.132, no.5, pp.1376–85.

Stengler-Wenzke, K., Müller, U., and Matthes-von-Cramon, G. (2003) 'Compulsive-obsessive disorder after severe head trauma: diagnosis and treatment', *Psychiatrische Praxis*, vol.30, no.1, p.37.

Taylor, S.F., Luan Phan, K., Decker, L.R., and Liberzon, I. (2003) 'Subjective rating of emotionally salient stimuli modulates neural activity', *Neuroimage*, vol.18, no.3, pp.650–9.

Taylor, W.D., MacFall, J.R., Payne, M.E., McQuoid, D.R., Provenzale, J.M., Steffens, D.C., and Ranga Rama Krishnan, K. (2004) 'Late-life depression and microstructural abnormalities in dorsolateral prefrontal cortex white matter', *American Journal of Psychiatry*, vol.161, no.7, pp.1293–6.

Tolin, D.F., Abramowitz, J.S., Przeworski, A., and Foa, E.B. (2002) 'Thought suppression in obsessive-compulsive disorder', *Behaviour Research and Therapy*, vol.40, no.11, pp.1255–74.

Toone, B.K. (1991) 'The psychoses of epilepsy', *Journal of the Royal Society of Medicine*, vol.84, no.8, p.457.

Trimble, M.R. and Cummings, J.L. (1981) 'Neuropsychiatric disturbances following brainstem lesions', *The British Journal of Psychiatry*, vol.138, no.1, pp.56–9.

Van Der Wee, N.J.A., Ramsey, N.F., Jansma, J.M., Denys, D.A., Van Megen, H.J.G.M., Westenberg, H.M.G., and Kahn, R.S. (2003) 'Spatial working memory deficits in obsessive compulsive disorder are associated with excessive engagement of the medial frontal cortex', *Neuroimage*, vol.20, no.4, pp.2271–80.

Veale, D.M., Sahakian, B.J., Owen, A.M., and Marks, I.M. (1996) 'Specific cognitive deficits in tests sensitive to frontal lobe dysfunction in obsessive-compulsive disorder', *Psychological Medicine*, vol.26, no.6, pp.1261–70.

Volle, E., Beato, R., Levy, R., and Dubois, B. (2002) 'Forced collectionism after orbitofrontal damage', *Neurology*, vol.58, no.3, pp.488–90.

Voon, V., Brezing, C., Gallea, C., Ameli, R., Roelofs, K., LaFrance, W.C., and Hallett, M. (2010a) 'Emotional stimuli and motor conversion disorder', *Brain*, vol.133, no.5, pp.1526–36.

Voon, V., Pessiglione, M., Brezing, C., Gallea, C., Fernandez, H.H., Dolan, R.J., and Hallett, M. (2010b) 'Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors', *Neuron*, vol.65, no.1, pp.135–42.

Vuilleumier, P., Richardson, M.P., Armony, J.L., Driver, J., and Dolan, R.J. (2004) 'Distant influences of amygdala lesion on visual cortical activation during emotional face processing', *Nature Neuroscience*, vol.7, no.11, pp.1271–8.

Weickert, T.W., Goldberg, T.E., Gold, J.M., Bigelow, L.B., Egan, M.F., and Weinberger, D.R. (2000) 'Cognitive impairments in patients with schizophrenia displaying preserved and compromised intellect', *Archives of General Psychiatry*, vol.57, no.9, pp.907–13.

Weinberger, D.R., Egan, M.F., Bertolino, A., Callicott, J.H., Mattay, V.S., Lipska, B.K., *et al.* (2001) 'Prefrontal neurons and the genetics of schizophrenia', *Biological Psychiatry*, vol.50, no.11, pp.825–44.

Weintraub, D., Koester, J., Potenza, M.N., Siderowf, A.D., Stacy, M., Voon, V., *et al.* (2010) 'Impulse control disorders in Parkinson disease: a cross-sectional study of 3090 patients', *Archives of Neurology*, vol.67, no.5, pp.589–95.

Wilson, B.A., Evans, J.J., Alderman, N., Burgess, P.W., and Emslie, H. (1997) 'Behavioural assessment of the dysexecutive syndrome', *Methodology of Frontal and Executive Function*, pp.239–50.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

PART 6

CHALLENGES, THEMES, AND ISSUES

Introduction

16 Cognition and emotion

Jenny Yiend, Bundy Mackintosh, and George Savulich

17 Autobiographical memory and the self

Clare Rathbone, Chris Moulin, Martin Conway, and Emily Holmes

18 Consciousness

Jackie Andrade

19 Cognitive modelling and cognitive architectures

Paul Mulholland and Stuart Watt

20 Theoretical issues in cognitive psychology

Tony Stone and Martin Davies

INTRODUCTION

The first four parts of this book help address the question 'What is cognitive psychology?', a question that we first took up in Chapter 1. Just as someone can answer the question 'What is a cat?' by pointing to examples of cats, so we can gain an understanding of cognitive psychology by considering different topics in cognition. As well as introducing substantive research areas, the chapters in the first four parts illustrated some of the breadth and depth of the cognitive approach, its assumptions and commitments, its methods, and its successes.

The emphasis changed somewhat in Part 5, which focused on contributions to cognitive psychology made by work within the neurosciences – neuropsychology, neuroimaging, and neuropsychiatry. The chapters in that part of the book discussed key features of work in those three areas, some of its methodological flavour, and some of the assumptions that are made in order to draw lessons for our understanding of cognition. Already you may feel that some of those assumptions could be contested, and that doing so would engender much debate. It is to some of these debates that we now begin to turn.

First, however, armed with an understanding of key examples of cognitive psychology, this final part of the book considers two different kinds of question. The first concerns how widely the cognitive approach may be successfully applied. Many research topics, such as emotion, the self, and consciousness, have been considered at various times to lie outside the purview of cognitive psychology. Yet cognitive psychologists have begun to apply their approach to these problems. The difficulties in so doing, as well as the successes that arise, are discussed in Chapters 16 to 18. What becomes clear in these chapters is that consciousness, an aspect of mind also discussed in Chapter 1, has become a theme for many chapters in this book. In Chapters 16 and 17, conscious experience plays an ever more important role until, in Chapter 18, it becomes the focus of our enquiry. The second type of question concerns the kinds of explanation that are pursued in cognitive psychology, a theme that has also developed in the preceding chapters. Chapter 19 focuses explicitly on cognitive modelling and cognitive architecture, the role that models play in theories of cognition, and criteria for their evaluation. As the authors demonstrate, a number of researchers have focused their attention on unifying theories of cognition, ones that are successful in accounting for evidence from all areas of cognitive psychology, including perception, attention, memory, language, and reasoning. In Chapter 20, where the authors adopt a more philosophical approach, we then turn to engage more fully with some of the ongoing debates about cognitive psychology and the nature of the mind. The chapter discusses several key theoretical issues within cognitive psychology that are helping to shape our understanding of the nature of cognition.

In Chapter 16, Jenny Yiend, Bundy Mackintosh, and George Savulich examine the relationship between cognition and emotion, aspects of the mind that Western thought has traditionally viewed as opposed to one another. Early in their chapter, the authors distinguish three components of emotions. There is a physiological or bodily response, such as heart rate (and the sweatiness of hands!) increasing at the sight of a loved one. There are emotional behaviours, in this example perhaps smiling broadly and, if the social context allows it, touching. And, third, there are emotional feelings, a glow of affection or, possibly, an intense rapture. The last of these components, the emotional feelings, are a large part of what many people think of when reflecting on their normal conscious experience. The 'affective tone' of a particular experiential episode, its positivity or negativity, is often one of its most salient features. Indeed, the authors of this chapter go on to look at evidence showing that people are better able to retain information that is congruent with their current mood than information that is not. That is, they remember happy information if it is presented to them while they are happy, and gloomy information presented to them when they are gloomy.

In addition, retrieval of information is best when mood at encoding matches mood at retrieval. You are better able to recall what you heard in a cheerful mood when you are again in a cheerful mood.

The relationship between cognition and emotion provides strong links between Chapters 16 and 17. Affective tone is again a prominent feature of autobiographical memories, which is the topic of Chapter 17. Clare Rathbone, Martin Conway, Emily Holmes, and Chris Moulin are interested in the relationship between autobiographical memory and the self, and their chapter has strong links also to the memory chapters in Part 3, with, for example, the episodic/semantic distinction of Chapter 8 making a reappearance. However, research into autobiographical memories cannot rely on the same kinds of laboratory-based, experimental methods used to investigate other types of memory – think of the difficulties of running a memory experiment where participants will be tested for recall decades later! Research in this area tends to rely on particular distinctive methods aimed at eliciting memories of actual episodes from our past, and the chapter activities invite you to take part in some of these procedures.

As with the emotional feelings discussed in Chapter 16, autobiographical memory also has a self-evident bearing on consciousness: much of the stream of consciousness features episodes of autobiographical memory or is derived from such episodes and from one's sense of self. The authors of Chapter 17 develop ideas about how autobiographical memory relates to what they call the 'working self'. This notion is modelled on the concept of working memory, the topic of Chapter 9, and which, because it refers to a person's current cognitive activity (what the person has in mind), is itself about consciousness. The working self is conceived as a hierarchy of interconnected goals, some but not all of which can enter consciousness. The chapter also examines the disruptive effects of affectively charged trauma memories that, in conditions such as post-traumatic stress syndrome, frequently intrude upon consciousness.

Consciousness has been a pervasive presence in many of the chapters in earlier parts, such as those on attention (why you are conscious of this rather than that), perception of objects and words (how you become conscious of this object or this word), recognition, and categorization (what you consciously experience this object *as* or that piece of discourse *as meaning*). And, as just indicated, the affective tone of experience is also a prominent feature of Chapters 16 and 17. In all these chapters, however, consciousness has played a supporting role whilst some other aspect of cognition has played the lead. In Chapter 18, however, consciousness comes out of the shadows to take centre stage. Jackie Andrade goes head-on with this most difficult and elusive of concepts. She outlines philosophical approaches to the vexed issue of consciousness and offers an analysis of the place of consciousness within cognitive psychology. She then goes on to examine a range of empirical cognitive research into the nature of consciousness. She starts by revisiting the topic of implicit memory that was discussed in Chapter 8, and goes on to extend this into a discussion of implicit learning. The evidence for learning without conscious awareness is difficult to interpret conclusively. Certainly, some implicit learning seems possible, but as far as the modern dream of effortless sleep learning is concerned, the news is entirely negative. Jackie Andrade looks into controlled versus automatic processing of information and then considers what neuropsychological evidence can tell us about the possible modularity of consciousness. Following this, she presents an analysis of what the function of consciousness might be, and this links back to ideas rehearsed in Chapter 16.

Chapters 16 to 18 thus present the results of applying the cognitive approach to difficult research questions that at one time were thought to be beyond the purview of cognitive psychology. As the authors themselves imply, cognitive psychology has succeeded in approaching these topics partly by identifying those aspects that might be amenable to the cognitive approach, and analysing them separately. It follows that there are aspects to emotion, the self, and consciousness – for example, emotional *feelings*, aspects of *identity*, *phenomenal* consciousness – that as yet remain recalcitrant, and may, for all we know, continue to elude our best attempts at explanation. Nevertheless, as the chapters indicate, the application of the cognitive approach has yielded some significant insights and numerous positive results, and interest in these research areas is likely only to increase.

As mentioned at the outset of this introduction, Chapters 19 and 20 then address a somewhat different set of questions. For example, both chapters focus upon the relationship between computation and cognition, picking up on a theme first introduced in Chapter 1. Paul Mulholland and Stuart Watt's Chapter 19 has the slightly daunting title of 'Cognitive modelling and cognitive architectures'. However, the authors do a fine job of rendering this austere-sounding topic comprehensible! The topic of cognitive modelling has been broached in many earlier chapters, such as Graham Hitch's Chapter 9 on working memory. In Chapter 19, the topic is illustrated and illuminated through an informative contrast between parallel distributed processing (PDP, or connectionist) models and rule-based systems. This leads the reader to the concept of a 'cognitive architecture', a concept that is needed to ensure a clear distinction between cognitive models and the computers on which they run. The notion of cognitive architecture is developed and elaborated by means of a careful overview of the ACT-R architecture. The authors go on to explain how ACT-R accounts for a range of phenomena found in the study of human memory. They then consider how ACT-R models the acquisition of skills of arithmetic and of mobile phone texting. Finally, they offer a comparison of rule-based and PDP architectures, followed by an analysis of the criteria against which the performance of a model should be judged.

In Chapter 20, Tony Stone and Martin Davies provide a theoretical overview of cognitive psychology with reference to three issues, each of which has been encountered a number of times in previous chapters. They revisit the contrast between rule-based and connectionist (or PDP) models that you will already have encountered in Chapters 1 and 19, and give a detailed evaluation of the two types of cognitive architecture within the context of the now famous past-tense debate. They demonstrate that the argument turns upon such complex matters as what it means for a rule to be explicitly or implicitly represented, and what it means for a representation to be compositional. The second theoretical issue Tony Stone and Martin Davies focus on concerns the modularity, or otherwise, of the mind. The notion of cognitive modules, or systems, has cropped up repeatedly since its first mention in Chapter 1 – for example in Chapter 8 on memory and Chapter 18 on consciousness. In Chapter 20, Fodor's theory of modularity is described and evaluated within the context of the philosophy of science. Modularity turns out to be another complex topic that continues to undergo theoretical refinement. The final theoretical issue considered in Chapter 20 is the relationship between cognitive psychology and the brain. Most if not all of the preceding chapters have included at least some appeal to neuropsychological evidence, but what is the status of such evidence? How does it bear upon cognitive theory? The discussion is related to Marr's three levels of explanation, a further connecting loop back to Chapter 1. Finally, the authors examine whether cognitive and neurobiological theories co-evolve or whether the former will eventually be reduced to the latter. Such theoretical questions as these, difficult as they are to resolve, help to provide an overarching conception of cognition: what it is, how we might model and theorize about it, and how such theories 'fit' in the wider picture of scientific enquiry.

CHAPTER 16

COGNITION AND EMOTION

Jenny Yiend, Bundy Mackintosh, and George Savulich

1 INTRODUCTION

This chapter is concerned not just with cognition but with how emotions influence, and are influenced by, cognitive processes. Emotions are such a familiar and fundamental aspect of everyday life, and it is often this very ability to experience and express emotion that is seen as a crucial distinction between the behaviour of humans and (possibly imaginary) high functioning computers or robots. As you have seen in previous chapters, it is possible for a computer to solve successfully many difficult tasks and so mimic human achievements. However, one very salient distinction between the performance of computer and human is that the computer won't show pleasure when reaching its goal or frustration if it fails, let alone empathy with its human operator. We shall consider whether this is really an advance later in the chapter.

Emotions seem such an important part of our lives that it might come as a surprise that despite the rapid development of psychology as a discrete scientific discipline since the mid-1800s, the study of emotion has largely taken a back seat. Why should this be so? One reason is undoubtedly the behaviourist legacy. Behaviourists such as John B. Watson (1878–1958) and Burrhus Frederick Skinner (1904–1990) recognized the need for scientific rigour and objective, verifiable measurement and were therefore exclusively concerned with the overt behaviours displayed by an organism – those that could be directly observed and measured. For behaviourists, reference to unseen mental processes was taboo. Their emphasis on objectivity and empiricism continues to be an important influence in cognitive psychology today. This historical bias for a long time deterred study of emotion, in which the main component – feelings – can only be accessed through introspection. Undoubtedly, another factor has been the attitude towards emotions often expressed in Western, European societies, among others. At least since Plato (375 BC), emotions have been viewed as impediments to rational thought. Darwin thought of them as childish or immature responses, a residual hangover from our evolutionary past that no longer had useful functions for the mature adult.

✱ ACTIVITY 16.1

Stop for a moment and reflect on one recent episode when you experienced emotion. Try to jot down three or more aspects that characterized this occasion as being emotional.

COMMENT

This is often a hard task, since one common feature of emotional situations is that it is sometimes difficult to put into words what is happening! However, you may have noted your feelings and perhaps what or who you considered was the cause of the emotion. Did you include a description of how you behaved or a change in body sensations?

One example could go something like this (note the different elements of this description): *One day when I was alone at home a special delivery van stopped at our house looking for an address nearby. I stepped outside to point out directions, the front door blew in the wind, and it locked behind me. 'No problem,' I thought, and went to the usual hiding place to retrieve the spare key. It was missing. Now I felt anger, frustration, and regret. Why wasn't the key in its usual place? I blamed others for not replacing it. Why had I been so stupid as to let the door slam? I swore and banged my fist. I went back to the door and rattled the handle. I was taking deep breaths and felt my heart beating fast with annoyance. I paced rapidly up and down whilst I tried to work out what to do next. . .*

However, the study of emotion and how it interacts with cognition has enjoyed a resurgence of interest. This is largely because of the development of objective, quantifiable ways of measuring the concomitants of emotion, such as psychophysiological techniques and brain imaging, and a shift in attitude towards the importance and function of emotions in everyday life.

1.1 Components of emotion

From the example in Activity 16.1 it is clear that there are different aspects to any emotional response. Traditionally, psychologists have identified at least three characteristics that are embodied within an emotional episode. These are:

- behaviours
- bodily responses (physiology)
- feelings.

1.1.1 Emotional behaviour and expression

Many of the behaviours associated with emotion will be familiar to you. Some simple examples include: laughing when you are happy; withdrawing from something you find disgusting; becoming agitated and raising your voice when you are angry; and becoming quiet, withdrawn, and slow when you are sad. Facial expressions that are characteristic of different emotions are also examples of behavioural responses. As these are all observable phenomena, they can be easier to study empirically than internal feelings. However, they cannot generally be used to infer emotions directly since, unlike most of the bodily (physiological) responses associated with emotion, emotional expressions can be brought under some degree of control. You can suppress your smile, make a special effort to appear cheerful when feeling sad, feign interest in order to be polite, and curb angry behaviour if it might jeopardize your well-being. Still more problematic for research, various cultures and social groups differ in their code of conduct with respect to emotional expressions. For example, in some cultures, such as in many Arabian countries, public grieving involves overt crying, moaning, and beating of the chest, which are seen as appropriate expressions of respect for the dead in assembled company, whereas in others, for instance in Japan, a polite smile and tight emotional control are expected.

Unlike the very individual feeling of emotion and the internal bodily response to it, emotional behaviours, including emotion expressions, are visible to others; that is, they can communicate (albeit imperfectly) the individual's emotional status. In any form of communication, understanding the mechanism fully requires researching both the recognition and the production of the appropriate signals (as discussed in Chapters 6 and 7). So far, much more research effort has been directed towards recognition than production of behavioural correlates of emotion. Facial expressions, rather than other emotion behaviour such as 'body language', have attracted most research attention.

1.1.2 Bodily responses

The bodily responses associated with emotions are the physiological reactions such as sweating when you feel anxious, or your heart racing when you feel agitated or excited. These reactions have been refined during evolution and are vital to survival. For example, if a lion attacks then there must be little delay before escape ('flight'), or maybe aggression ('fight'), begins. Typically an animal's body will respond to this kind of threat by diverting the blood flow away from less vital regions, such as the gut (digesting lunch is suddenly less important than ensuring you don't become someone else's). The extra blood, and therefore energy, is supplied to the major muscle blocks and the brain (rapid processing of information is needed or energy could be wasted running in the wrong direction). The contents of the blood are altered, boosting 'fuel' in the form of blood glucose and cholesterol, and increasing clotting agent (which serves to stem blood loss in case of wounds), and so on. This physiological reaction stands us in good stead when physical action is required to ensure survival.

Many of these bodily responses are controlled by the autonomic nervous system (ANS) (see Figure 16.1), a

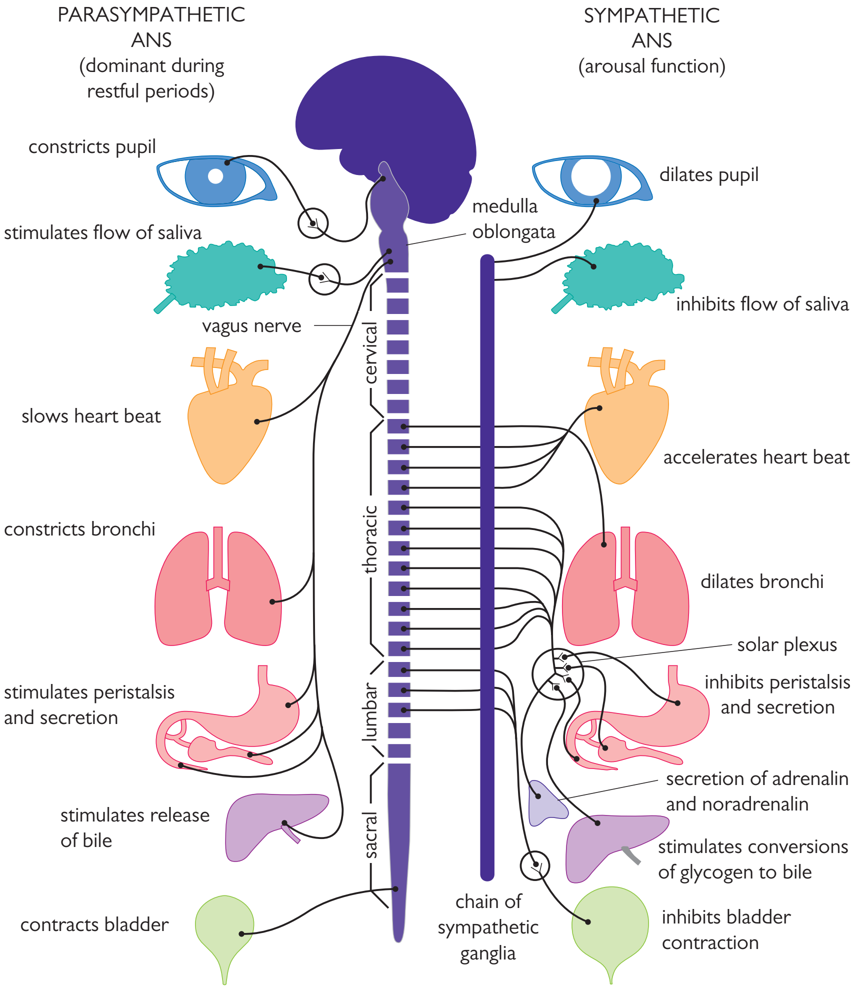

PARASYMPATHETIC
ANS
(dominant during
restful periods)
SYMPATHETIC
ANS
(arousal function)
constricts pupil
dilates pupil
medulla
oblongata
stimulates flow of saliva
inhibits flow of saliva
vagus nerve
cervical
slows heart beat
accelerates heart beat
thoracic
constricts bronchi
dilates bronchi
solar plexus
stimulates peristalsis
and secretion
inhibits peristalsis
and secretion
lumbar
secretion of adrenalin
and noradrenalin
stimulates release
of bile
stimulates conversions
of glycogen to bile
sacral
contracts bladder
chain of
sympathetic
ganglia
inhibits bladder
contraction

FIGURE 16.1 The autonomic nervous system (ANS) and physiological responses associated with emotions, showing sympathetic and parasympathetic sections. *Source: Reber, 1995, p.76*

network of nerve fibres throughout the body that transmits signals to the various organs, muscles, and glands. The ANS is divided into two sections. The **sympathetic ANS** produces effects associated with arousal. These include secretion of the hormone adrenalin from a gland near the kidneys. Adrenalin release initiates and enhances sympathetic activity leading to changes such as accelerating heart rate, vasoconstriction (constriction of the blood vessels), increased respiration (breathing) rate and depth, and reduced gastrointestinal (gut)

BOX 16.1 METHODS Measuring emotion using psychophysiology

Most emotional states tend to lead to increased arousal and therefore produce corresponding physiological signs. Psychologists have been able to devise ways of measuring this physiological change precisely. For example, by applying a tiny electrical current across the fingers we can measure the electrical resistance of the skin. This changes according to minute differences in the amount of sweat produced and so provides a physiological measure known as either GSR (galvanic skin response) or SC (skin conductance), correlating with changes in arousal (recall the use of GSR responses to detect the influence of 'shocked' words in the non-attended messages discussed in Chapter 2, Section 1.3). The measurement of changes in GSR has also been used as a 'lie detector', picking up individuals' emotional responses arising during deception.

Electromyography (EMG), another psychophysiological measure, records muscle tension and activity using small electrodes placed on the skin (usually below the lower eyelid; see Figure 16.2). EMG is most commonly used to measure the **startle response**, an eye-blink reflex that protects the eye from a potential threat. 'Startle' or 'blink' magnitude is typically measured using a 50-millisecond burst of white noise that causes small changes in muscle contraction surrounding the eye. Magnitude of startle has been shown to be faster and of greater strength while the individual is exposed to an aversive stimulus.

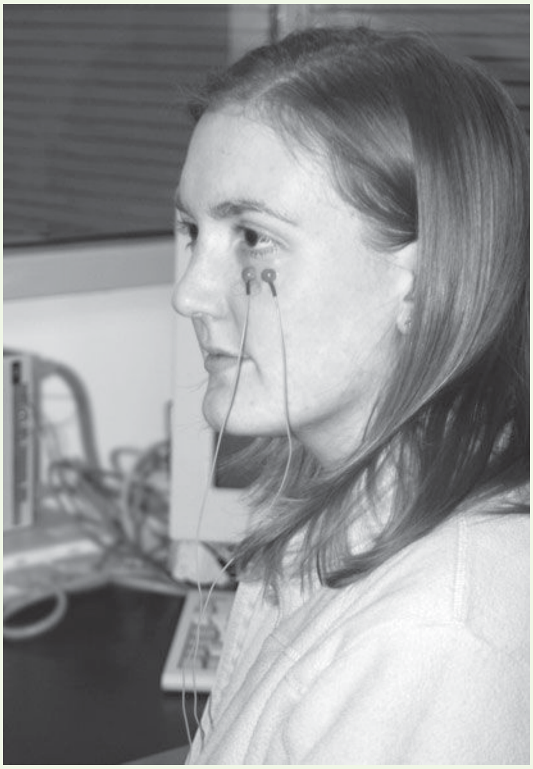


FIGURE 16.2 Measuring EMG: electrodes are placed under the eye and measure 'startle' or 'blink' magnitude, which is the muscle contraction produced when you blink or are surprised, for example in response to a sudden loud noise.

Other common psychophysiological measures include heart rate (usually measured as beats per minute using a simple transducer that converts the movement produced by the pulse into electrical energy), blood pressure (the pressure in the arteries as blood is circulated), cortisol levels in the blood or saliva (related to adrenalin production), respiration rate (usually measured as the number of breaths per minute), and surface skin temperature (related to dilation or constriction of the blood vessels).

activity. The intricate pattern of changes in hormone levels, breathing, redirection of blood flow and pressure and changes in its constituents, and the many other changes occurring under stress all prepare the body for physical exertion (the 'flight or fight' response described above). This is a physiological pattern fine-tuned by evolution and shared by other mammals, with only relatively minor details differing between species. In contrast, the **parasympathetic ANS** tends to dominate during periods of rest, having broadly opposing effects on the body.

Box 16.1 discusses techniques for measuring physiological responses that investigate this component of emotion.

1.1.3 Feeling emotions

Feelings are private and subjective. They are, by definition, states of experiential awareness. Around the world, humans can usually report a wide range of different states or feelings, from anger to fear to love, which can be recognized and understood by those around them. Within psychology the feeling component of emotion is inextricably bound up with notions of conscious awareness and the subjective self (see Chapters 17 and 18). Emotion researchers are often interested in whether the stimuli or tasks that they use elicit positive or negative feelings, and if so to what degree. They certainly acknowledge that feelings co-occur with the other markers of emotion, but otherwise cognitive psychologists have concentrated less on feeling states than on exploration of the cognitive processing associated with emotions and emotional information.

Now try Activity 16.2.

One issue that researchers face when trying to study any of the components of emotion described above is how to elicit realistic effects in the laboratory. In real life, emotions are usually stronger than those seen in the laboratory, so we have to be cautious when generalizing from the results of any given study. There are at least two reasons why it is difficult to study the behaviours, physiology, and feelings associated with *strong* emotions:

1. It is sometimes unethical to induce strong emotions, for example strongly negative ones, in a laboratory setting.
2. Strong emotions are hard to elicit in a predictable fashion and take a while to die down, making laboratory study impractical.

For these reasons, you will find throughout this chapter that the study of cognition and emotion is largely confined to consideration of relatively mild emotional states and the processing of mildly emotional information. However, the information gained from such work provides useful insights about possible cognitive processing when emotion is more extreme.

ACTIVITY 16.2

Thinking back to what you've learned so far about cognition, would you say that emotions may also have a cognitive component, as well as the components we have just discussed (behaviour, bodily responses, and feelings)? If so, we would be able to study the cognitive side of emotions using the techniques and paradigms familiar to cognitive scientists.

COMMENT

Looking only at the three components of emotion discussed above, you might well answer 'no' with regard to bodily responses and feelings. The behavioural component of emotions includes face perception, which will be familiar from Chapter 4, though in emotion research it is the emotional expression and not identity that is of interest. However, as we proceed you will see that many of the tools developed by cognitive scientists can be widely adapted to the study of emotions. Not only do cognitive processes interact with emotions, but many psychologists believe that they are an integral part of producing the emotions themselves.

SUMMARY OF SECTION 1

- Although the study of emotions from a cognitive point of view has historically been neglected, the advent of new techniques and new ideas about the significance and function of emotions has brought a resurgence of interest in the study of emotion and how it interacts with cognition.
- There are thought to be three main components of emotions:
  - emotional behaviour and expression (e.g. emotional facial expressions)
  - bodily responses (e.g. galvanic skin response)
  - feelings.

2 DIFFERENT EMOTIONS

You should now have some idea of what is meant by the term 'emotion' in psychology. Next we shall consider how one might classify and explain the huge variety of different emotions that individuals typically report. Psychologists usually take one of two approaches to dealing with the task of accounting for different emotional experiences. Some refer to **basic emotions**, a categorical set thought to be of biological origin. Others take a **dimensional view**, which considers the full range of emotional experience and the dimensions used to describe them.

2.1 Basic emotions

One approach has been to assume that underlying the richness of emotion experience there are a small number of discrete emotions – ones considered to be the most fundamental or important. This idea is analogous to the processing of colour by the visual system, where the whole range and subtlety of our colour experience is achieved through stimulation of just three different types of cones in the retina. Likewise, it is argued that different combinations of 'basic emotions' can produce all the other emotions. For example, a mixture of joy and acceptance produces friendliness according to Plutchik (1991), a prominent basic emotions theorist.

At the core of basic emotion research is the concept that emotions are a product of evolution. This belief maintains that a categorical set of emotions are biologically based and are therefore characteristic of humans and many other species. Paul Ekman, one of the most influential psychologists in this tradition, first used the term **emotion family** (1992) to describe sets of emotions related to each other when considering 'theme' (characteristics unique to a particular emotion family) and 'variations' (differences within each theme). Sad expressions, for example, can be distinguished by narrowing of the eyes and crinkling of the forehead that vary depending on the intensity of the sadness experienced. However, themes characteristic of sadness are fundamentally different from themes characteristic of other emotions (such as tightly-pressed lips expressing anger or upward-turned corners of the mouth expressing happiness). The basic emotions approach argues that these are natural phenomena, and that emotion families should be approached from biological and evolutionary perspectives. Izard (2007) proposes that basic emotions are motivated by neural, bodily, and feeling components that activate evolved neurobiological substrates. These substrates change with development and are thought to influence the feeling states of basic emotions. Despite growing evidence from brain imaging techniques that show specific associations between patterns of neural activation and some of the basic emotions, there are several distinct challenges. One is to provide evidence for the existence of a small number of discrete emotion states; another is to decide how many emotions should be called basic and which ones they are. The idea of basic emotions has considerable general support but few agree exactly on the appropriate number and type of emotions that should be included. This point is illustrated by Table 16.1.

Despite these widely differing views there are five emotions, sometimes called **the Big Five**, that appear to represent a broad consensus among basic emotion psychologists. These are anger, fear, sadness, disgust, and happiness. Ekman, building on research described by Darwin in *The Expression of the Emotions in Man and Animals* (1998, first published in 1872), has collected a formidable body of information from cross-cultural studies to support the fundamental status and importance of these five emotions. He was impressed by the observation that wherever he travelled people displayed broadly similar emotions, and that he had no difficulty in interpreting them despite language barriers. For more systematic research his main method was to show pictures of facial expressions, such as those in Figure 16.3(a), and determine whether people from different cultures consistently selected the same emotion label to describe each one.

TABLE 16.1 Basic emotion theorists and the emotions they propose

Emotion theorist	Fundamental emotion
Arnold	Anger, aversion, courage, dejection, desire, despair, fear, hate, hope, love, sadness
Ekman, Friesen, and Ellsworth	Anger, disgust, fear, joy, sadness, surprise
Frijda	Desire, happiness, interest, surprise, wonder, sorrow
Gray	Rage and terror, anxiety, joy
Izard	Anger, contempt, disgust, distress, fear, guilt, interest, joy, shame, surprise
James	Fear, grief, love, rage
McDougall	Anger, disgust, elation, fear, subjection, tender-emotion, wonder
Mowrer	Pain, pleasure
Oatley and Johnson-Laird	Anger, disgust, anxiety, happiness, sadness
Panksepp	Expectancy, fear, rage, panic
Plutchik	Acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise
Tomkins	Anger, interest, contempt, disgust, distress, fear, joy, shame, surprise
Watson	Fear, love, rage
Weiner and Graham	Happiness, sadness

Source: Power and Dalgleish, 1997

Figure 16.3(b) shows some typical results for six emotions – the Big Five plus surprise. Although there is some variation, particularly within isolated non-literate cultures, there is always agreement above what would be expected if people were just guessing (the 'chance' level, shown by the white bars). Ekman also provided evidence for basic emotions in the production as well as in the recognition of expressions. He visited a visually isolated non-literate group in New Guinea (people who had not previously met or seen pictures of anyone from outside their own cultural group) and asked them to show him what their face would look like if they were sad, happy, and so on. He then took videos of their expressions and played them back to American students, who had to decide which emotion was being displayed by the New Guineans. The American judges had no problem in identifying the different emotions according to the (translated) labels to which the New Guineans had been responding, which further supports the notion of pancultural or universal emotions (Ekman *et al.*, 1969; Ekman, 1999, provides a review of all his work).

In support of the basic emotions approach, Ekman provides extensive evidence from cross-cultural work such as ratings of spontaneous displays of emotion across different cultures. More convincingly, he has used objective measurements of facial behaviour (how much different parts of the face move) and compared these across cultures and countries – for instance, by testing participants from Japan and the USA (Ekman, 1973). His studies have also extended to infants from different cultures (Ekman and Oster, 1979). You may

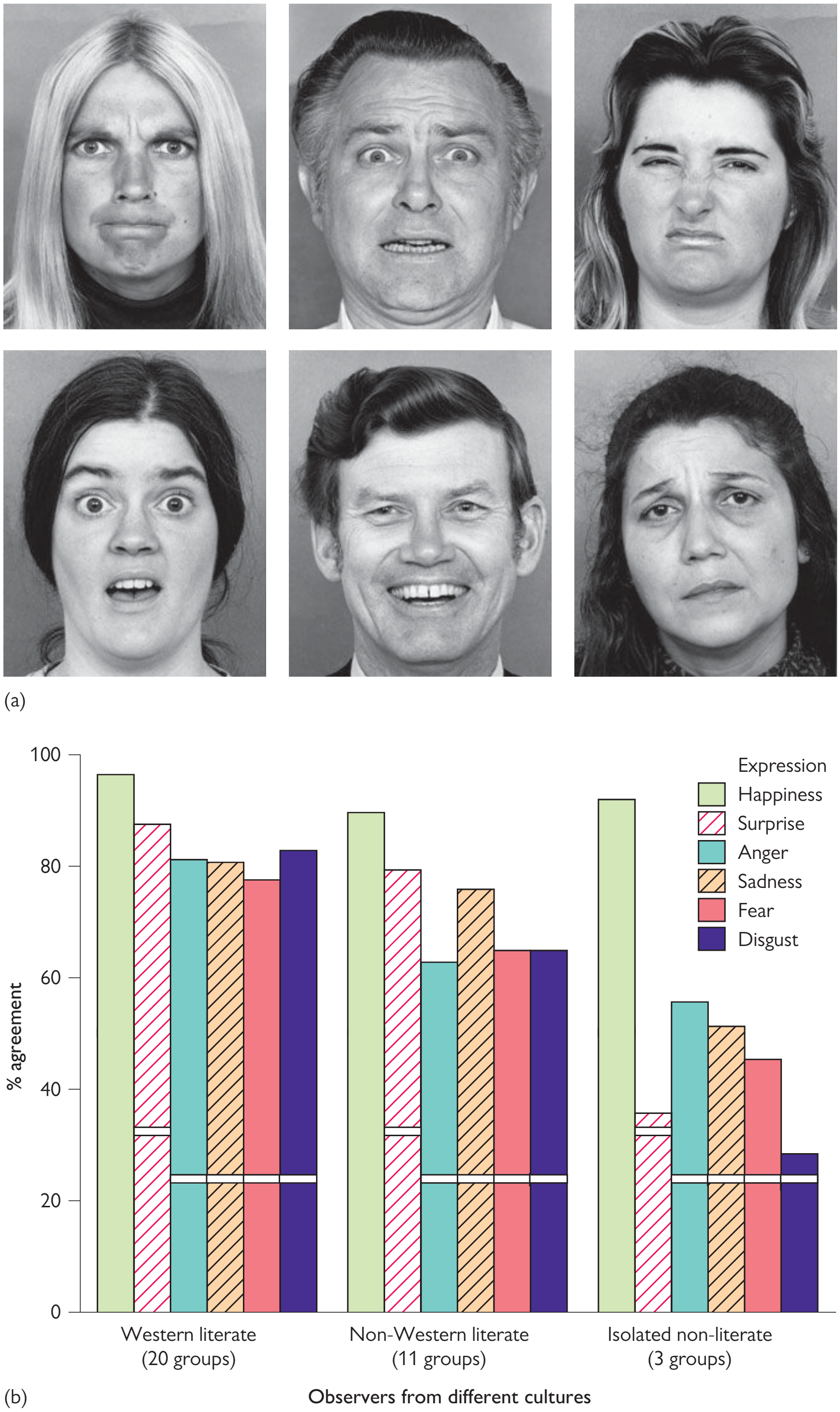


FIGURE 16.3 (a) Some of the photos of facial expressions used by Ekman, showing (left to right from top) anger, fear, disgust, surprise, happiness, and sadness. (b) Results from cross-cultural studies showing differences in recognizing facial expressions of six emotions. *Source: (a) Ekman and Friesen, 2003; (b) Rosenzweig et al., 1999, Figure 15.3, p.414*

wonder why Ekman chose to look particularly at infants' facial expressions rather than those of adults. If a characteristic or ability is present in infants, who have had little opportunity to be influenced by their culture or upbringing, then that is additional evidence for that characteristic being largely genetic rather than learned. The spontaneous facial expressions of blind children (Medicus *et al.*, 1994; Eibl Eibesfeldt, 1988) also support the idea that there may be basic emotions, and further that they may have biological rather than social origins. Other theorists, such as Plutchik and Frijda (Plutchik and Landau, 1973; Frijda, 2001; Frijda and Tcherkassof, 1997), rely not only on facial expressions but on whole body movements – what is often called 'body language'.

Ekman's assumption about the inheritance of emotion is shared by many others promoting the notion of basic emotions. For these researchers it follows that such emotions arise from sub-cortical brain mechanisms that we still share with many other species (e.g. Panksepp, 1989; Panksepp *et al.*, 1991; LeDoux, 1989). Debate is still active concerning whether and which emotions are basic, or whether it is clusters of related emotions that should be considered together. However, many believe that the development of the brain systems underlying something approximating to the basic emotions seems to have arisen far back in our evolutionary past before the separation of mammals, reptiles, and birds. Box 16.2 discusses how technology for imaging the brain has begun to shed light on the relationship between specific emotions and particular brain structures.

BOX 16.2 METHODS Imaging emotions in the brain

PET and fMRI

Although techniques for imaging the structure of the brain (taking pictures without the need to make any actual physical intrusion) have been used in medicine for several decades, the ability to study changes in activation associated with brain function is a more recent and rapidly growing technique (see Chapter 14). Two common techniques are PET (positron emission tomography) and fMRI (functional magnetic resonance imaging). PET involves injecting the participant with very slightly radioactive water, which then travels around the body including the brain, emitting its radiation as it goes. Participants perform an experiment usually designed to test particular hypotheses about which brain areas are involved in the task(s) concerned. Because the most active areas of the brain will draw the most blood, these areas will also be emitting the most radioactivity. This is measured using special gamma ray (the energy component of radioactivity) detecting equipment, and thanks to complex software this can be translated into brain images, which look similar to those shown in Figures 16.4 and 16.5. fMRI also produces similar-looking images of brain activation involved in a task, but using a different method. A strong magnetic field is applied across the brain that aligns certain particles in the blood ('deoxygenated haemoglobin' molecules) in the same direction (similar to the way iron filings line up towards a magnet). When this field is removed these particles 'precess', or move back again, and in doing so each particle emits a discrete 'package' of energy that is detected by specialist equipment. The more active an area of the brain is, the more of these particles it has, and consequently the more energy is emitted during precession.

Traditionally, PET and fMRI studies mapped brain functioning through use of **block designs** that present successive trials of conditions that are always grouped together. A simple block design, for example, would instruct participants to look at a set of words (Block A) followed by a set of pictures (Block B). Average neural activation patterns across each block would then be compared to test for differences in brain activation while the participant was viewing the two sets. Although block designs demonstrate strong statistical power and are useful in detecting brain activation changes between alternating conditions, they do not take into account stimulus predictability or allow assessment of brain changes occurring within individual trials. It is these latter effects that are often of most interest to psychologists. The more recent onset of **event-related designs** addresses these issues by randomized presentation of trials and averaging of neural activity across trials of the same type. Importantly, event-related designs allow researchers to analyse their data based on the participant's performance (e.g. correct responses only) and eliminate the potential confounding effects of **habituation** (a *decreased* neural response to the stimulus presented).

Overall, block and event-related designs are two approaches used for evidencing associations between neural activation and distinct emotions, and have already revealed some very interesting results.

Neural substrates and emotion

Many studies using the brain imaging techniques such as those discussed above have now shown that a structure called the **amygdala** is involved in the processing of all types of emotion and is particularly strongly activated in response to fear stimuli. Similarly, two areas are implicated in recognizing disgust: the insula, an area of cortex (the convoluted outer layer of the brain) and the basal ganglia (an evolutionarily old area in the brain stem). This has been corroborated by data from a patient who has damage to these areas and is particularly poor at recognizing disgust in others (Calder *et al*., 2000).

Figure 16.4 shows the areas of the brain where different studies have reported activation resulting from either the processing of fearful faces (green squares) or learning about fear (red circles). As the figure shows, the activation triggered by processing fearful faces tends to involve the left amygdala, whereas learning about fear seems to produce more bilateral (i.e. left *and* right) activation.

Figure 16.5 shows the areas where studies have found brain responses to disgust. As with Figure 16.4, the two images depict different slices through the brain. The insula activations are shown in blue

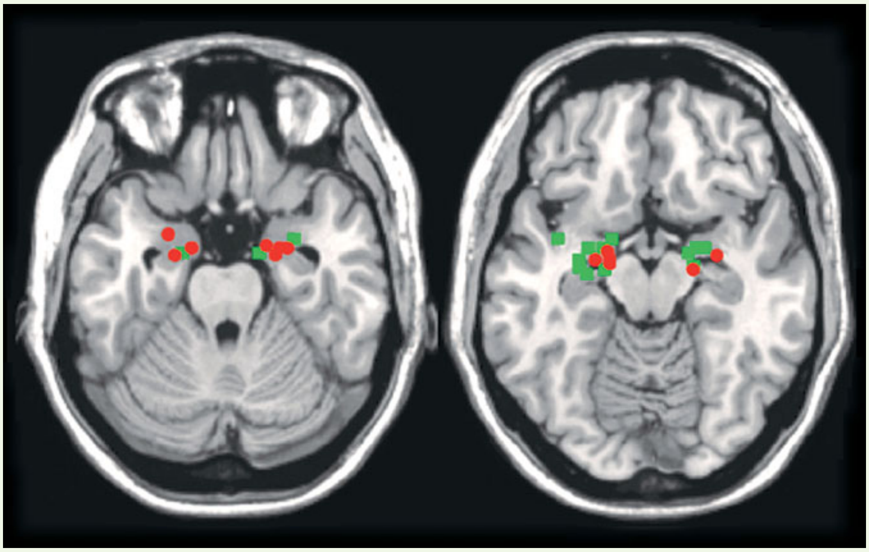


FIGURE 16.4 Amygdala activations in response to the processing of fearful faces (green squares) or learning about fear (red circles). The image on the left is a horizontal 'slice' through the brain, with the eyes at the top end and the back of the head at the bottom. The image on the right is the same sort of slice but taken higher up, more towards the top of the head. There is a tendency for the activation triggered by processing fearful faces to involve the left amygdala, whereas learning about fear seems to produce more bilateral activation. *Source: Calder et al., 2001, Figure 3, p.357*

and basal ganglia activations in red. The basal ganglia signals are mainly in the right hemisphere, whereas the insula signals are more evenly distributed across both hemispheres.

Another feature of these new brain-imaging techniques is that as well as the cognitive processing of emotion described above, they can give us an objective measure of the 'feelings' side of emotion (see Section 1.1.3). As we said earlier, feelings have been notoriously hard to study in psychology because the only way to measure them was to rely on people's subjective self-reports of their own internal state – the much-scorned 'introspection'. Now, though, we can investigate how brain activity changes according to the strength and nature of our feelings, and this is a possibility that is only just starting to be exploited.

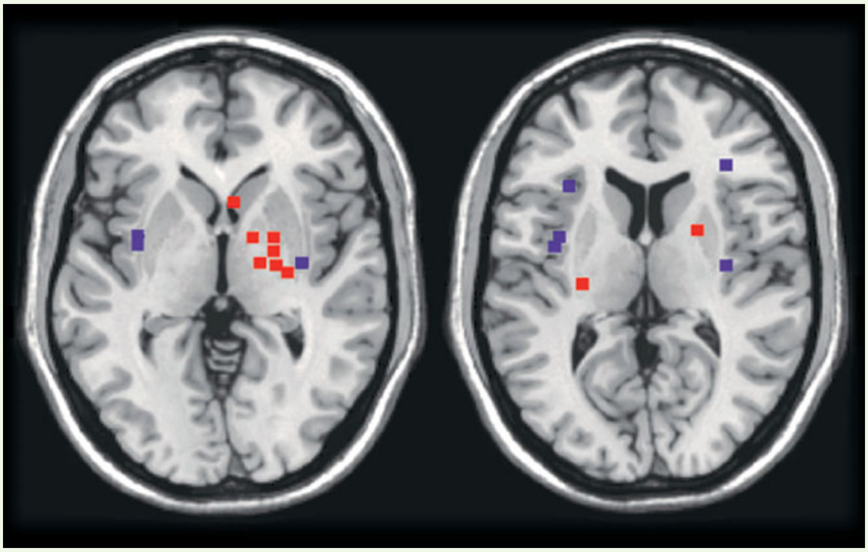


FIGURE 16.5 Insula and basal ganglia activations in response to disgust. The two images depict different slices through the brain. The insula activations are shown in blue and basal ganglia activations in red. The basal ganglia signals are mainly in the right hemisphere, whereas the insula signals are more evenly distributed across the two hemispheres. *Source: Calder et al., 2001, Figure 4, p.357*

2.2 Verbal labels

The fact that very similar verbal labels for emotions are used across widely differing languages and cultures is sometimes used as evidence in support of the existence of a discrete set of basic emotions corresponding to those labels. Scherer and colleagues (2001) have studied vocalizations – the sounds we use to communicate – to address whether the recognition of basic emotions is universal. 'Simulated' vocal expressions were obtained by asking German actors to record multiple nonsensical phrases in an emotional tone. Participants from nine countries then listened to these recordings and were instructed to identify which emotion they thought was being expressed. An overall accuracy of 66 per cent was found for correct identification of anger, sadness, fear, joy, and neutral portrayals. Interestingly, comparisons between countries revealed that accuracy decreased as the native language of the participant resembled German less, and that patterns of errors were highly consistent. These data support the idea that emotional vocalizations are recognized across cultures but that differences in language have an effect. In other work by Scherer and colleagues (e.g. Scherer and Wallbott, 1994a, b; Wallbott and Scherer, 1988), verbal labels were compared for seven emotions (anger, fear, sadness, joy, disgust, shame, and guilt) by translating the English terms into the languages of 37 countries. If all languages include words to describe the so-called basic emotions, and these emotions can be recognized cross-culturally, then that gives reason for believing in the universality of the concepts for the basic emotions.

What about all the other emotion words; where do they fit into the idea of basic emotions? Scherer and others introduced the idea of 'modal emotions'; that is, the idea that a number of these other emotion words may cluster together under a common 'theme', and that the specific clustering of emotion words betrays the underlying emotion concepts of the individual (recall discussions about concepts in Chapter 5). To complicate matters, different languages and cultures do seem to differ in the number and categorization of their emotion terms. It is not surprising to find that the range of situations that trigger emotions varies across cultures, but in addition, different emotions are either elaborated or downgraded in emphasis. There appears to be a set of universals – for instance, loss of a loved one leading to sadness, and attack to fear or anger – as well as a multitude of cultural specifics, such as whether looking directly at a woman's face evokes sensations of polite interaction, flattery, or insult. Whilst debate continues about whether there are a small number of basic emotions and whether these are necessarily inherited, it is clear that there are many cultural differences in emotions. Thus, there are cross-cultural differences in:

- the number and type of complex emotions
- the triggers for many emotions
- the socially acceptable rules governing display of emotions in certain contexts.

2.3 The dimensional approach

The concept of 'basic emotions' is not without challenges. Theorists such as Barrett (2006) note that even though humans describe their experiences using a discrete set of emotions, psychologists have yet to agree on the exact criteria that define them. Ortony and Turner (1990) have asked why, if basic emotions are so basic, there is so much disagreement about which count as basic, with some contenders (e.g. interest and desire) sometimes not even being considered as emotions at all. An alternative dimensional approach, as the name implies, assumes that the full range of emotional experience can be explained by identifying a few key dimensions. If there are only two key dimensions, then all emotions could be identified as being located in a two-dimensional space specifying the relative contribution provided by each of the two dimensions. An example of this approach is shown in Figure 16.6.

The 'affect grid' in Figure 16.6 is taken from work by Peter Lang and colleagues (Bradley *et al.*, 1999), who concentrated on studying our physiological responses to emotional material. There are two dimensions, **arousal** and **valence** (valence refers to the positive/pleasant or negative/unpleasant qualities of something). The figure shows people's ratings of how 'aroused' and how 'positive or negative' they feel about a variety of different pictures. Other, separate dimensions, such as 'dominance', have also been proposed, producing a more complex three-dimensional space (dominance reflects a quality related to how dominated vs in control the participant feels when considering this emotion or emotional information). This dimensional approach circumvents some of the problems associated with basic emotions. It has the advantage of suggesting how the different emotions relate to one another and makes it easy to understand how different languages could have developed different words to describe subtly different mixes of emotion experience. However, some emotions appear to combine attributes that the dimensional model assumes should be at opposite ends of a single continuum. For instance, nostalgia seems to combine attributes of positive valence (the positive value of past experiences) with negative valence (sadness or regret at their passing). Similarly, the excitement of extreme sports or roller-coaster rides combines pleasure (positive valence) with fear (negative valence) to create the characteristic exhilaration and excitement. Furthermore, it is still

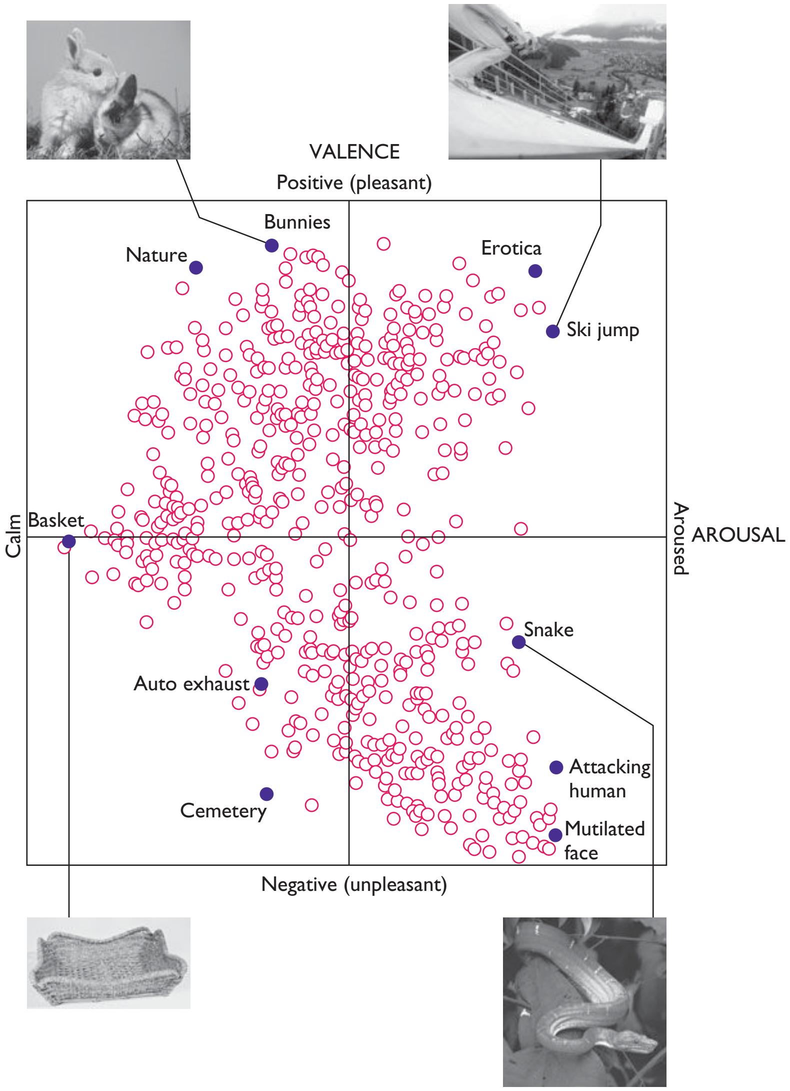


FIGURE 16.6 The affect grid and example pictures (the ratings on the grid are for pictures similar to those shown). *Source: Dawson et al., 1999, Figure 8.2, p.161*

✱ ACTIVITY 16.3

You might like to consider how you would map the different discrete emotions onto the affect grid. Where would you place sadness, contentment, fear, and excitement, for example?

COMMENT

You have probably opted for bottom left, top left, bottom right, and top right for sadness, contentment, fear, and excitement respectively. Sadness, for example, could be considered as fairly unpleasant with little excitement or energy. But notice too that the grid allows for a lot more variation between items. You may also notice that the distribution has a 'C' shape to it. It appears that there are plenty of things that we consider to be neutral (neither pleasant nor unpleasant) and not particularly arousing, but very few arousing neutral items! Putting it another way, if something is arousing we tend to find it either really good or really bad (or as already noted, maybe both together).

necessary to determine the dimensions, and how many should be used, and to decide how these relate (if at all) to the evidence suggesting the existence of basic emotions (see Activity 16.3).

SUMMARY OF SECTION 2

- Some psychologists classify different emotions by identifying discrete or basic emotions.
- The Big Five basic emotions are widely recognized by many psychologists and there is reasonable evidence to support this classification.
- The use of similar verbal labels, and production and recognition of emotional expressions across different cultures, further supports the notion of basic emotions.
- An alternative approach is to use a small number of continuously varying dimensions to describe the range of emotional experience.

3 THE FUNCTION OF EMOTIONS

Emotions and emotional responses to events could surely not have evolved unless they served a useful purpose, but what purpose or purposes might these be (see Activity 16.4)?

3.1 Emotions alter goals

One influential modern theory of the function of emotions is that of Oatley and Johnson-Laird (1987). They have proposed an evolutionary account of emotions that suggests the role of emotion is to signal that ongoing behaviour should be interrupted to take account of a conflicting goal (see Table 16.2). They argue that humans have many different motivations and goals. Events will happen that require setting or resetting of priorities amongst these goals, such as giving up the goal of planting next summer's food crop in favour of running away from an attacking lion. For example, sadness caused by bereavement is not maladaptive in their framework, but is seen as having the function of initiating readjustment of life goals that included the lost one. When the relationship was close, this period of reassessing or reforming goals could be lengthy.

Central to Oatley and Johnson-Laird's theory is the notion of cognitive readjustment to emotional events.

✱ ACTIVITY 16.4

Can you think of any aspect of emotions already touched upon in this chapter that might bestow a useful advantage on animals (including humans)?

COMMENT

Think (or look) back to Section 1.1.2 on physiological responses to emotion. You will recall that, in response to a frightening event, rapid physiological changes take place that prepare the body for 'fight or flight'. Undoubtedly the rapid mobilization of the body's resources in this way provides a potentially life-saving advantage.

TABLE 16.2 Summary of emotions and their associated goals according to Oatley and Johnson-Laird

Emotion	Juncture of current plan	Behaviour/ response
Happiness	Sub-goals being achieved	Continue with plan, modifying as necessary
Sadness	Failure of major plan or loss of active goal	Do nothing/ search for new plan
Anxiety	Self-preservation goal threatened	Stop and attend vigilantly to environment and/or escape
Anger	Active plan frustrated	Try harder, and/ or aggress
Disgust	Gustatory goal frustrated	Reject substance and/or withdraw

Source: Oatley and Jenkins, 1996, Table 9.1, p.256

However, the exact mechanisms are not spelled out. Unlike the physiological changes that we discuss next, the mechanism behind changes in cognitive processing in response to emotions is much less well understood. Later in the chapter (Sections 4.2 and 4.3) we touch on some aspects of attentional deployment and memory in emotion, but there is still a considerable shortfall in our understanding of how internal emotional status influences cognition and how processing of emotional information is prioritized and influences cognitive function.

3.2 Emotions mobilize physiological resources

It is relatively easy to see how the physiological changes involved in emotions are part and parcel of the need to readjust goals, sometimes with great rapidity. It is vital that if your life is threatened, then your body is ready to respond in the best possible way to ensure your survival. In Section 1.1.2 we described some of the bodily responses associated with emotions such as fear and gave an example of how these assist in ensuring survival. The physiological reactions described there stand us in good stead when physical action is required. Often in modern life, however, an emotional threat requires not more physical exertion, but less. One example is the threat of a pending examination, which requires long hours sitting still at a desk to revise rather than any physical exertion of the body. Similarly, most of us have experienced the fear of a near accident while driving, but all our bodies actually need to do to avoid the danger is perform minimal, albeit rapid, movements of the hands on the steering wheel and the feet on the brakes.

Does this mean that many emotional reactions, especially fear responses, no longer have useful functions? No. There are still many occasions when rapid physical responses avert death or injury. Even before an exam, when you won't be fighting or fleeing (even if you feel you'd like to), the increased adrenalin and physiological arousal will provide an energizing effect that can improve performance, if maintained at an optimal level. However, too much anxiety impairs performance as the anxiety itself interferes with cognitive function and the physiological reaction makes it hard to relax and sit still. At the other end of the spectrum, not enough arousal – in other words boredom or disinterest – also impairs performance.

This finding, expressed formally by Yerkes and Dodson (1908), is known as the **Yerkes–Dodson law**, and is shown in Figure 16.7. Notice that, for an easier task, higher levels of arousal are needed to attain the optimal level of performance compared with a hard task. The Yerkes–Dodson law seems reasonable and suggests that appropriate levels of emotion can indeed be useful, while both too much and too little emotional arousal can put an organism at a disadvantage.

3.3 Emotional expressions as communication

What function might emotional expressions fulfil and can we see evidence of them in other animals? Charles Darwin formalized the evolutionary view of the emotions in one of his later works entitled *The Expression of the Emotions in Man and Animals* (1998, first published in 1872). As well as acknowledging the more

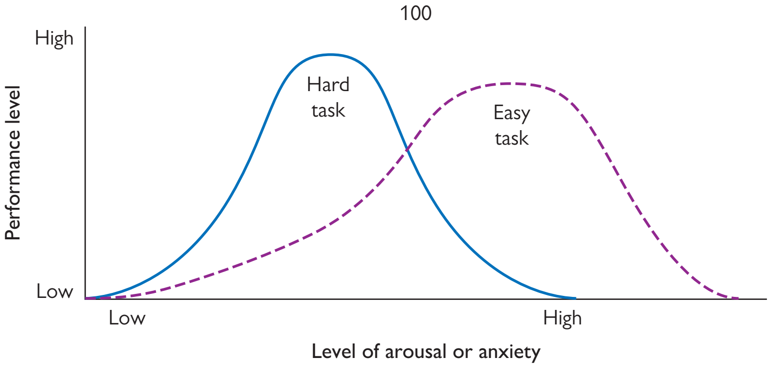


FIGURE 16.7 Yerkes–Dodson law. *Source: Eysenck and Keane, 1996, Figure 18.8, p.454*

obvious evolutionary advantage of physiological changes during emotions (such as preparation for fight or flight), he highlighted how expressions of emotion serve to communicate the emotional status of an animal to others of their species (so-called conspecifics). However, it has to be said that Darwin felt that emotions and emotion expressions in humans were no longer functional, but merely a relic from our evolutionary past – much as our appendix is seen to be superfluous to digestion. He drew parallels between the expressions of animals and their functions, such as the snarl of a dog communicating a readiness to bite, and the sneer of a human, which presumably has the same origin but no longer sends the same message (normally!) (see Figure 16.8).

Darwin might have been partially correct in his feeling that emotion expressions no longer have the function in humans that they do for other animals. We have already considered that display rules for expressions differ from culture to culture. For this difference to occur, then of course we need to control the expression displayed (at least to some extent). This also means that humans are capable of deceiving with their expressions – we can lie about our emotional feelings. Thus, emotional expressions serve multiple functions for humans: they can be honest signals of emotional status, as in other animals, or they can be part of the impression management, polite interaction, or social manipulation of the sender.

3.4 Emotions as information

To illustrate the idea of emotions as information, let us return to the Capgras delusion first mentioned in Chapter 4 (Section 6, Box 4.2). To remind you of the

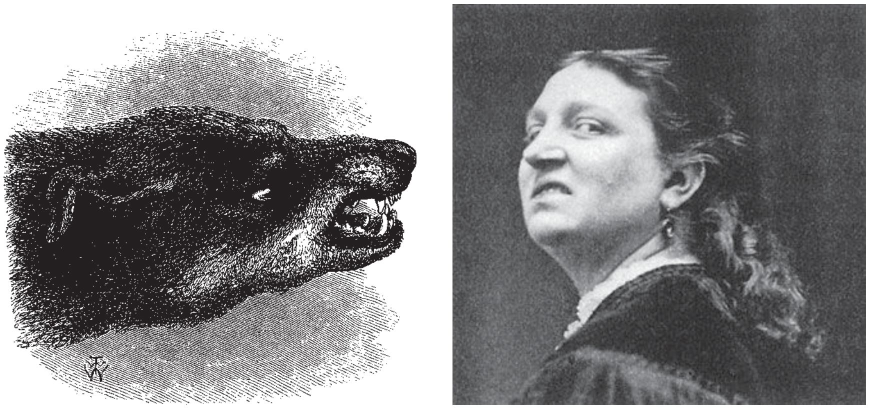


FIGURE 16.8 Darwin's comparison between the sneer of a woman and the snarl of a dog. *Source: Darwin, 1998, Figures 14 and 22, pp.117 and 246*

delusion (or syndrome) we will illustrate it with the case of an individual whom we will call Alan. Alan was in a car accident with his wife and sustained an injury to his head. His wife, Christine, was also injured and taken to hospital but later recovered. However, Alan refused to believe that Christine was still alive. He recognized her face but remained convinced that this was not really her, but a sinister impostor. Remember that in this rare syndrome the sufferer believes that a family member, or someone close, has been replaced by aliens or impostors. In such cases it is believed that although facial recognition is intact, a parallel system for registering the emotional meaning of the face has been damaged. Indeed, when Alan's palms were tested for SC changes (skin conductance – see Box 16.1) when viewing pictures of Christine's face, these responses were absent. SC changes, signalling an emotional response, would normally occur for any of us if viewing either emotional expressions or the face of someone we knew. Their absence signals a lack of emotional resonance; no sense of affiliation. It would appear that Alan's brain interpreted this lack of physiological feedback as evidence that this was not someone close. However, since the perceptual qualities of the face matched those of his wife, it must be someone who looked just like Christine, an impostor, a frightening and distressing situation for all.

The idea that emotions provide information to guide decision making is fundamental to the theories of Damasio (1996). His views are best explained by describing the task most associated with him, the so-called **Iowa gambling task**. You are given four decks of playing cards and asked to select from one and turn the card over. For reasons that are not explained, you are either rewarded or fined as a result of your selection. Your task is to attempt to maximize your winnings. After playing for a while you are likely to find yourself making more selections from two out of the four decks, but you probably won't be able to say exactly why. This task is arranged so that two decks (the 'good decks') give less spectacular wins but also less punishing losses. Choosing from these two over a period of time achieves a modest gain. The other two 'bad decks' sometimes deliver large wins, but also large losses resulting in an overall loss on average. The rule is not hard and fast and so is not generally very obvious as you play. Playing the 'good decks' is the best strategy; you win a little and lose a little, but overall your winnings start to add up. Damasio has developed a theory to explain how people come to operate this strategy successfully, and make other similar decisions in life, when they are acting on hunches rather than full understanding.

According to Damasio, the emotional responses to winning and losing produce physiological changes that he calls **somatic markers**. Over time, through a process of conditioning, the decks come to evoke different physiological responses in the player, essentially representing the accumulated positive emotions of wins together with negative emotions of losses. After extended experience, as the player considers making a selection from each deck, the physiological response conditioned to this deck will be initiated and this acts as a marker capable of guiding choice. Damasio suggested that somatic markers represent the 'gut feelings' that we often use to guide our decisions, even though we may never become consciously aware of why we have a gut feeling about a particular choice. The function of emotion, for Damasio, is therefore centred around information and future actions.

3.5 What is the function of emotional feelings?

Although we have only touched on the topic, it is relatively easy to propose evolutionary advantages conferred by emotional behaviours and physiological responses to emotions. The same is not true of the function of emotional feelings. Take fear, the example we have used the most. The physiological response to a fearful situation can provide the physical resources to escape danger or stand and fight. Our behaviour, including expressions, functions to communicate our fear to others. But why do we need to experience the unpleasant *feeling* of fear, or anticipated fear (anxiety) when we expect a frightening experience?

Feelings are part of our conscious experience. The functions of consciousness, as you will consider in

Chapter 18, are by no means uncontroversial, but one facet that is fairly regularly acknowledged is the notion that consciousness is necessary for performing new tasks or trying to override habits without relying simply on mechanisms of conditioning. This would also apply to learning new responses to an emotional situation, such as when soldiers continue to advance into battle despite a strong urge to flee, or in overriding habits, such as when suppressing the tendency to respond in anger at a socially inappropriate moment. However, whilst these examples invoke the need for consciousness, they still do not explain the necessity to *feel* the emotions of fear or anger. As the psychology of emotion continues to develop, future theories and research are likely to give us greater insight into the possible function of the feelings associated with emotion.

SUMMARY OF SECTION 3

- Following Darwin, many psychologists believe that emotions have evolutionary functions, including the mobilization of physiological resources, which remain valid today.
- Oatley and Johnson-Laird maintain that the purpose of emotions is to interrupt current behaviour in order to change priorities and goals in the light of new information.
- Damasio has formulated a somatic marker account of the function of emotions, in which their primary purpose is to provide information, via bodily feedback, which guides future decision making.
- The functions of 'feeling' emotion are still speculative.

4 EMOTION INFLUENCES COGNITION

4.1 Some important concepts

Before we start to discuss how cognition and emotion interact with each other, there are some important terms that you need to become familiar with. The first of these is the distinction between state and trait emotion.

4.1.1 State and trait emotion

State emotion (also called mood or affect) refers to how you feel right now. As you will be aware, this can change from minute to minute, day to day. State emotion is a very transient and variable entity. It is a construct that allows us to acknowledge the fact that momentary feelings may be quite different from the way an individual usually feels. Although state emotions are usually measured by self-report (asking participants to introspect and describe how they feel), they also relate directly to the behaviours and physiology discussed above and can be measured in the same way.

In contrast, **trait emotion** refers to more stable personality characteristics or 'what kind of person' you are. For example, some individuals may be prone to angry outbursts, or have a tendency to worry about things, or be optimistic and always looking on the bright side. Psychologists have directed much effort into trying to capture theoretically these ideas about stable personality characteristics. Thus, traits are theoretical constructs relating to aspects that are more enduring and characteristic of a person, and describe how one person may differ from others. Some common traits that have been proposed and are frequently measured (again by self-report) include: anxiety; depression; social desirability (how much you adapt your behaviour in order to gain the approval of others); anger; impulsivity; and emotional sensitivity.

A trait tends to make a person more prone to experiencing the associated mood state. For example, a high trait anxious individual will tend to feel more anxious for more of the time than a low trait anxious person. This is why certain traits, like anxiety or depression (sadness), are useful to psychologists interested in emotions – they are a more permanent indicator of who tends to have more or less of the relevant state emotion.

4.1.2 Processing vs manifestation of emotion

Another important distinction is between the *manifestation of emotion* itself and *the processing of emotional material*. The manifestation of emotion is exactly what we were discussing in Section 1. Thus, by 'manifestation' we mean both the experience of emotion, the feeling state, and the expression of that experience through bodily changes and behaviours. This is also often known as the 'hot' component to emotion. In contrast, the 'cold' component is the processing of emotional material but without emotion being actually experienced. This isn't always an easy distinction to make. It is a bit like the difference between describing an emotional event in a detached way (relating a series of facts) compared with describing it in emotional terms. Obviously the two types of process regularly co-occur – the memory of the facts of an event often brings back the feelings as well – and in this case the manifestation/processing distinction may seem blurred. However, in psychology we often use stimuli such as words or pictures as a way of studying how we process emotional material, although these stimuli rarely elicit a strong experience of emotion in participants. It is important to grasp, then, that studying cognitive processing in emotion can be quite distinct from studying the manifestation of emotion.

Already you may have realized that the processing of emotional material is our first example of an interaction between cognition and emotion. In a typical experiment one might present participants with lists of negative emotional words (e.g. cancer, attack, evil) mixed with neutral words (e.g. number, unusual, round) and ask for later recall in a surprise memory test. The emotional aspect in the task is the valence (pleasantness/unpleasantness) of the words, which is the independent variable. The cognitive measure (dependent variable) is how many words of each type are recalled in the memory recall test. You might be interested to know that while most people will remember more positive than negative words (a very common 'positive bias' in emotion processing), individuals with clinical depression tend to remember more of the negative words. Box 16.3 in Section 4.2.1 discusses this further.

As we mentioned above, things can become more complicated when hot and cold emotions occur simultaneously. In the psychology of cognition and emotion we are interested not just in how people process emotional material, but also in how this processing is affected by emotional states and traits. For example, does the processing of sad words change when someone is actually feeling sad at the time? Similarly, we might want to know whether people who are vulnerable to anxiety (i.e. high on trait anxiety) process threatening words any differently from those who are not. These more complex questions are what cognition and emotion psychologists are mostly concerned with. Section 4.2 delves deeper into these issues.

4.1.3 Cognitive biases

A final term that is important to this section is **cognitive bias**, defined as the *selective processing* of pathology-congruent information. In other words, cognitive processes are biased if an individual selectively prioritizes information that reinforces a pathological state. This then leads to a self-perpetuating cycle of cause and effect, in which the information processed is thought to contribute to the cause and maintenance of a psychopathology. There is some evidence already consistent with this. Laboratory-based experiments have developed 'training' programmes in an effort to modify the cognitive biases potentially contributing to emotional disorders. These programmes use **cognitive** bias modification (CBM) procedures that experimentally train participants to select for further processing information that is either congruent or incongruent with their pathology (e.g. Wilson *et al.*, 2006). Such induction manipulations are thought to address the same underlying cognitive mechanism as cognitive behavioural therapies (CBT). However, there is no direct empirical evidence that this assumption is true. Evidence for the effectiveness of training studies and their general implications will be reviewed for the cognitive processes of memory, attention, and semantic interpretation discussed next.

4.2 Memory

We start our examination of the interactions between cognition and emotion by considering the ways in which emotional states affect memory for emotional material.

4.2.1 Mood congruent memory

What happens to memory processes when the content of material being encoded matches the mood state of the participant doing that encoding? For example, if you are feeling sad and then happen to watch a sad film, how does this influence your later memory of the film? This scenario could produce an example of mood congruent memory (MCM). Bower and colleagues' classic experiments (e.g. Bower, 1981) sparked a great deal of interest in this phenomenon. In a typical example, participants are put in either a happy or sad mood by hypnosis and then read both a happy and a sad story (you may like to consider the ethical implications of such a study). Participants are then given a surprise recall test to see how much of each story they can recall. The results are shown in Figure 16.9.

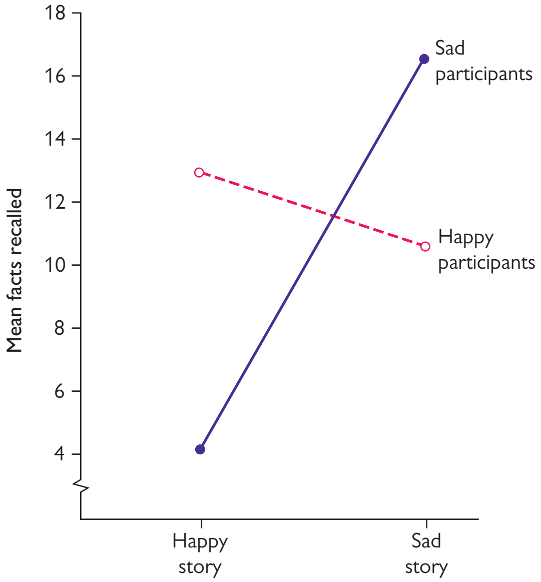


FIGURE 16.9 Results from Bower's (1981) mood congruent memory experiment. *Source: Eysenck and Keane, 1996, Figure 18.4, p.446*

As you can see, more was recalled from the story that matched the mood of the participant as they were reading. For example, sad participants recalled more facts about the sad story than about the happy one. The phenomenon of mood congruent memory has proved very robust. It has also sparked a whole field of research into the effects of emotional disorders on cognitive processing, such as the relationship between clinical depression and memory processes (see Activity 16.5).

Interestingly, work on mood congruency has alerted us to the finding that even 'normal' individuals, in no particular mood, seem to have a positive and potentially adaptive bias towards memory for positive information. Some suggest that this helps us to keep a positive outlook on life, in the face of all the problems

✱ ACTIVITY 16.5

You might like to think about what you would predict if you tested a clinically depressed person on memory for negative and neutral information, bearing in mind that one of the hallmarks of clinical depression is chronic low, or sad, mood. After you have considered this, look at Box 16.3.

BOX 16.3

Typically, individuals with clinical depression and those who are not diagnosed but still report feeling constantly low in mood (sub-clinical depression) all show mood congruent memory (MCM) effects, sometimes called a 'bias', for negative material. Many different types of experiments have been used to verify this finding, using positive and negative word lists, self-descriptive adjectives, sentences, and whole scripts (Matt *et al*., 1992 offer a meta-analysis). The effect appears to be stronger when participants are aware of the relationship between their mood and the material, and also, not surprisingly, when the negative nature of the material is stronger (e.g. 'evil' vs 'bad'). The bias also includes recall of autobiographical memories (see Chapter 17). Although this method might seem inconclusive (maybe depressed people really have had more negative experiences anyway), experiments using mood induction really do suggest that mood affects the valence of the personal memories that are brought to mind.

These findings are of more than just theoretical interest. It has been suggested that MCM may contribute to keeping someone in a depressed mood and that if we change this cognitive processing bias, then that might help the mood to lift. Teasdale (1988) has developed this idea, as shown in Figure 16.10. The suggestion is that patients' bias towards recalling more negative mood congruent information means that their world will seem to contain more negative things than is really the case. This in turn will make them feel even more depressed. You can see that a vicious cycle could be set up, where the memory bias contributes to the mood, which enhances the memory bias and so on. Teasdale and others have spent a lifetime of research trying to devise methods of breaking this cycle and coming up with new cognitive treatments for depression, such as a procedure called mindfulness-based cognitive therapy (Segal *et al*., 2002).

Joormann and colleagues (2009) have used suppression training in individuals with major depressive disorder (MDD) to induce forgetting of negative information. Depressed and non-depressed participants first learned to associate neutral cue words with positive and negative target words. Participants in the training condition were then instructed to suppress negative target words by learning to associate positive or negative 'substitute' words with the original cue words. It was found that depressed participants successfully forgot negative words in both the positive and negative substitute conditions but did not in the unaided (no training) condition. This study demonstrates that suppression training reduces recall of negative information in depressed individuals and that *thought substitutes* are useful tools for inducing cognitive change. Suppression training is therefore an appropriate method for modifying memory congruent bias and will be important to further investigation of the effects of forgetting on depressed mood.

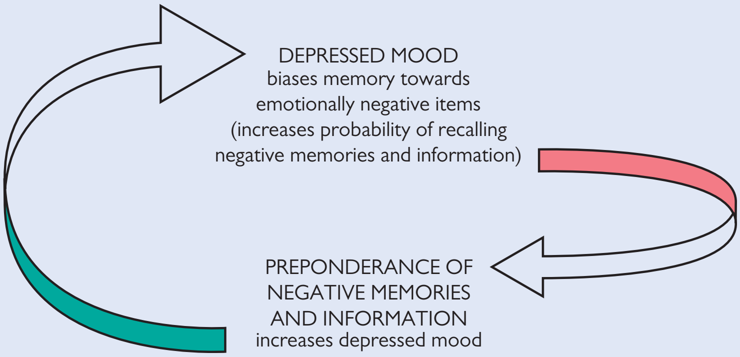


FIGURE 16.10 It is thought that the mood congruent memory effect contributes to a vicious cycle in which depressed mood enhances the accessibility of negative memories. In turn, having more negative memories in mind is likely to exacerbate depressed mood. *Source: based on Teasdale, 1988*

it throws at us. It is as if we are 'looking at the world through rose-coloured glasses'!

4.2.2 Mood dependent memory

Mood dependent memory (MDM), or mood state dependent recall, is a well-known but controversial phenomenon. It can be seen as a specific case of the influence of context on memory that was described in Chapter 8. The idea is that your memory for a particular stimulus or event will be better if there is a match between your mood at the time you experienced it and your mood when you try to recall it. For example, imagine you have a heated argument with a friend. Mood dependent memory would suggest that you will remember more of what was actually said if you are in an angry state again than if you are not.

In the laboratory this hypothesis has been tested using the following type of experiment. Participants are put into particular moods (mood induction) by one of several techniques such as hypnosis, listening to appropriate music, or reading appropriate passages of text. Then they are asked to learn a list of arbitrary, neutral words while in the induced mood. Participants are later put back into either the same or a different mood and asked to free recall the words ('remember as many as you can'; no cues or prompts are given). If this second induced mood matches the one they were in when they learned the first list, then recall should be higher.

A classic experiment of this type is that by Bower (1981) who used happy and sad mood induction by hypnosis. Figure 16.11 shows some typical results from the experiment. In this design participants learned two lists of words, list A and list B, one after the other, but only recall for the first list, list A, was tested. As usual, participants were put into either a happy or sad mood before learning took place, one mood for each list. Thus those who learned list A in a happy mood then learned list B in a sad mood, and vice versa. Then, after both lists had been learned, they were tested on their recall for the first list only. The mood of participants during the test (using a third mood induction) either matched or contrasted with the mood at the time of learning list A. So for some participants mood at recall matched mood at learning (points 1 and 2 in Figure 16.11), whereas for others mood at recall was different from mood at learning (points 3 and 4 in Figure 16.11). You should be able to see from the figure that when learning and test moods were the same, participants were indeed better at remembering list A, compared with participants who tried to recall the same list in a different mood from the one they were in when they learned it.

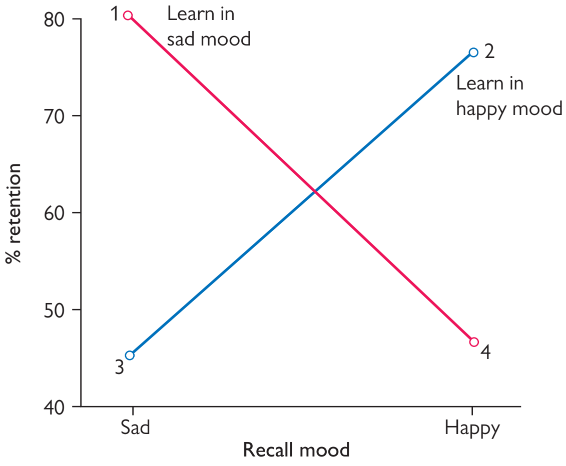


FIGURE 16.11 Percentage retention of words according to the match between learning mood (happy or sad) and recall mood. *Source: Based on Bower, 1981, Figure 2, p.132*

Perhaps you are wondering what was the point of the second list B? The reason for using two lists was simply that learning list B in a contrasting mood acted as an interference task, which made the experiment more sensitive to the beneficial or detrimental effects of the mood manipulations.

Bower (1981) went on to propose an influential **semantic network** theory to explain these mood and memory effects. The theory is shown in Figure 16.12. Bower suggested that emotions could be represented as nodes in a network, having numerous connections to related semantic items (words, concepts, etc.), other emotion nodes, and outputs such as behaviour and autonomic responses. Material such as memories and knowledge is stored in the network and may be connected to some emotion nodes. Nodes become activated by external or internal stimuli and when this happens that activation selectively spreads across the network via the links to other units, a bit like ripples across a pond. Notice that some connections are inhibitory, so that

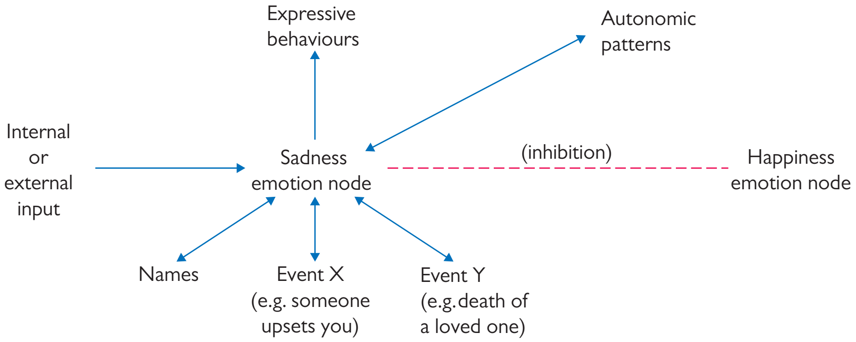


FIGURE 16.12 An example semantic network theory of emotion. *Source: Power and Dalgleish, 1997, Figure 3.2, p.71*

activation of the sadness node, for example, would suppress any activation in the opposite happiness node. When nodes are activated above a certain threshold, then the content of those nodes enters conscious awareness leading to the corresponding feelings and thoughts.

You can perhaps start to see how this theory fits with the results of mood dependent memory experiments. When participants learn a word list in one mood, links are created between the relevant emotion node and the memory representations of those words. Thus when participants try to recall the same words this can be made easier if they are in the same mood thanks to the spreading activation from the associated emotion node. Conversely, in a different mood there will be no advantage from such activation, and indeed it is assumed that inhibition of the word representations from the incongruent emotion node might result.

Since Bower's original experiment there have been many attempts to replicate his finding, but these have met with varied success. It seems that mood dependent memory is not a very robust effect. It is much influenced by factors such as the strength of the mood state that is induced and the nature of the items to be recalled (e.g. recall using real-life autobiographical events produces better results). However, in a review of the work on mood dependent memory Eich and Metcalfe (1989) concluded that the phenomenon itself was genuine, and that the problems lay with the methods used to detect and measure it. There is no doubt that Bower's findings and theory have been remarkably fruitful in their influence on the thinking and direction of emotion research.

Before we move on, it is worth stopping to think what the key difference is between mood dependent memory experiments and those we discussed in Section 4.2.1 under the heading mood congruent memory. Here we have been concerned merely with the effect of mood on recall, irrespective of what it was that was actually being remembered. With mood congruent effects, however, there is always a match – or congruity – between the emotional material being recalled and the mood of the individual when encoding that material. Congruity means a match between mood at encoding and material being encoded (see Activity 16.6); dependency refers to a match between mood at encoding and mood at retrieval.

✱ ACTIVITY 16.6

Can you think of examples where congruent or incongruent stimuli (rather than mood states) might influence cognition?

COMMENT

It has regularly been shown that an individual's performance is influenced by whether or not two separate aspects of a situation are matched or not. For example, in tasks demonstrating the flanker effect (see Chapter 2, Section 3.3) you may remember that performance depends on how closely matched the targets and the distractors are. The Stroop effect, discussed in Section 4.3 below, is another such example.

4.3 Attention

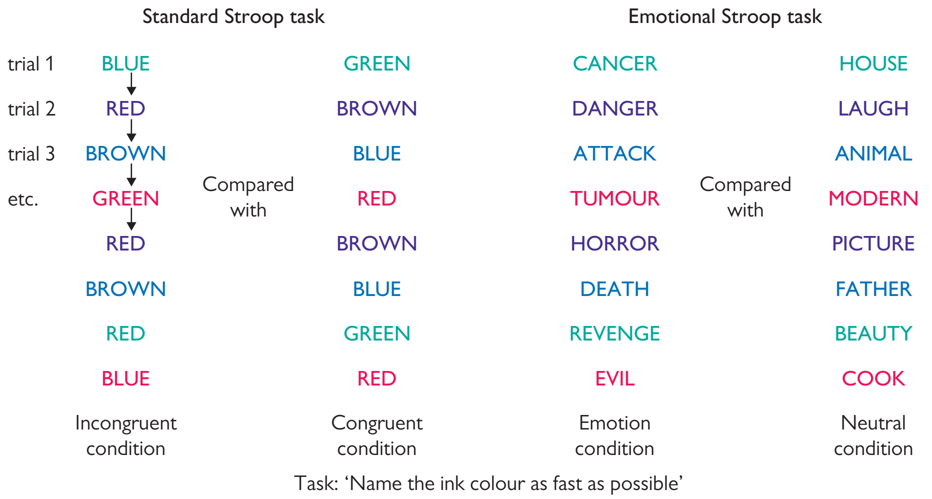
In the same way that memory for emotional material can be biased in a direction consistent with one's mood, so can attention. A classic example of this is the 'emotional Stroop'.

In the standard **Stroop task** (Stroop, 1935) (see Chapter 2, Section 3.3, Box 2.2), participants are asked to name out loud, as fast as they can, the colour of the ink in which colour words are written. When the ink colour is different from the meaning of the word itself (e.g. 'blue' written in red ink) participants are slowed down compared with stimuli where the word meaning and ink colour are matched. The effect arises because of the different amounts of interference between congruent ink colour and word meaning compared with a competing or incongruent ink colour and word meaning. The emotional Stroop differs in that, instead of colour words, emotional and neutral words are used, still printed in different colours. Examples of both types of Stroop task are shown in Figure 16.13.

When the emotional Stroop is given, for example, to high-trait anxious individuals, then the interference from the anxiety-relevant words is usually greater than that from the neutral words, compared with the same difference when observed in non-anxious individuals. As performance on the Stroop task is generally taken to be a measure of attention towards the word meanings (although the precise mechanisms behind the effect are still not fully understood), then this is an example of an *anxiety-related attentional bias*.

In an attempt to demonstrate more clearly the nature of this attentional bias, MacLeod *et al.* (1986) published a now classic paper using an innovative new method of testing attention allocation. Their design, now known as the **dot probe** or 'attentional probe' task, is shown in Figure 16.14.

The task is to respond as rapidly as possible to the presentation of a dot (termed a 'probe' because it is probing where attention is located). This is, therefore, a **reaction time (RT)** task. On some trials (catch trials) there is no dot, to make sure participants are really looking for it and not just responding as soon as the words disappear. As you can see, before the dot a pair of word stimuli are displayed, one threatening and one neutral. If a participant is consistently faster to find the dot whenever it appears where the threatening item was, then we can reasonably assume that they


Standard Stroop task
Emotional Stroop task
trial 1
trial 2
trial 3
etc.
BLUE
RED
BROWN
GREEN
RED
BROWN
RED
BLUE
Compared with
GREEN
BROWN
BLUE
RED
BROWN
BLUE
GREEN
RED
CANCER
DANGER
ATTACK
TUMOUR
HORROR
DEATH
REVENGE
EVIL
Compared with
HOUSE
LAUGH
ANIMAL
MODERN
PICTURE
FATHER
BEAUTY
COOK
Incongruent condition
Congruent condition
Emotion condition
Neutral condition

Task: 'Name the ink colour as fast as possible'

Standard result
Incongruent slowed compared with congruent condition (because word meaning interferes with colour naming on incongruent trials).

Emotional result
Anxious participants: emotion condition slowed compared with neutral condition.
Non-anxious participants: no difference between emotion condition and neutral condition.

FIGURE 16.13 The standard and emotional Stroop.

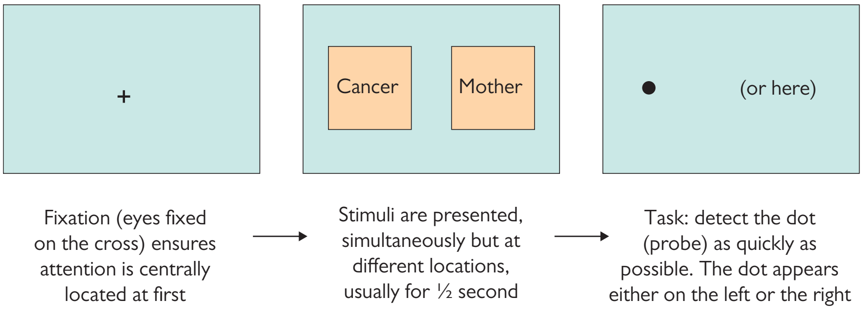

+
Cancer
Mother
(or here)
Fixation (eyes fixed on the cross) ensures attention is centrally located at first
Stimuli are presented, simultaneously but at different locations, usually for ½ second
Task: detect the dot (probe) as quickly as possible. The dot appears either on the left or the right

FIGURE 16.14 The dot probe task.

must have been attending to that item rather than to the neutral item. The original results of MacLeod *et al.* (1986) are shown in Figure 16.15.

The figure shows that control (not anxious) participants were just slightly faster when probes appeared in the neutral rather than the threat areas of the display (another example of the normal 'positive bias'). Anxious patients were the other way round – faster for probes appearing where threat words had been than for probes appearing where neutral words had been. This strongly suggested that anxious individuals allocate their attention to threat words rather than to neutral words, whereas controls do not. Thus, consistent with the emotional Stroop results, MacLeod *et al.* found an attentional bias for threat in their anxious patients. These results sparked over a decade of continuing research into this so-called attentional bias for threat. We now know that the bias is seen with many different types of material including words, pictures, and faces, but is most prominent when the material matches the current concerns of the individual. For example, snake phobics will show a stronger attentional bias towards pictures of snakes than towards pictures of snarling dogs. This type of bias, with suitable materials, has been shown with patients suffering a variety of anxiety disorders, such as those with phobias, generalized anxiety, and post-traumatic stress disorder. It is also apparent, although less reliably so, in non-clinical individuals who have high state anxiety or high trait anxiety, or both.

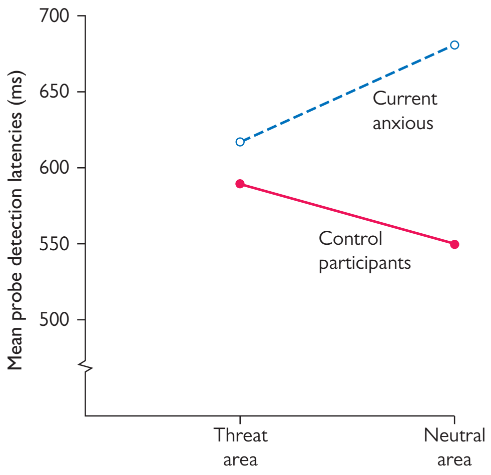


FIGURE 16.15 Results from MacLeod *et al.* (1986).
Source: Eysenck and Keane, 1996, Figure 18.6, p.450

Look again at Figure 16.10, where it is suggested that depression could be exacerbated by a vicious cycle of mood congruent memories contributing to sad mood. Could a similar mechanism be involved in attentional processing? Mathews (1990) proposed that just such a vicious cycle could operate to maintain anxious mood and attention to threat. Imagine that your anxiety makes you pick out and pay more attention to potential threats in the environment. This bias may well make it seem as if your surroundings are full of threats and this would, unsurprisingly, make you feel more anxious, which would perpetuate your attentional bias and so on. You would end up in a hyper-vigilant state, anxious about, and on the look out for, threats of relevance to you. This idea that anxious people are constantly in a vigilant, checking mode is popular in current theorizing about clinical anxiety.

How could you test whether an attentional bias might cause anxiety or be caused by anxiety (or both), and can anything be done to reverse its effects? Mathews and MacLeod (2002) set themselves such a task by devising methods of directly inducing a positive or negative bias in non-anxious volunteers using specialized training procedures, and then assessing its effects on anxiety levels. Possessing an induced positive bias reduced anxiety levels when exposed to a moderately stressful situation just after training, whilst a negative bias increased anxiety levels. These results certainly confirm that the attentional bias has a causal effect on anxiety levels, and interestingly that training procedures have been found that directly modify the bias, and can thereby reduce (or increase) anxiety. Such training procedures have since expanded to the development of cognitive bias modification programmes that are specifically designed to induce attentional bias toward non-threatening stimuli (Cognitive Bias Modification for Attention, or CBM-A). Amir and colleagues (2009) implemented an eight-session CBMA programme, designed to reduce attentional bias toward threat and decrease anxiety, in a population characterized by generalized anxiety disorder (GAD). This was achieved using the dot probe task by training participants to attend to non-threatening words on 66 per cent of the trials and then testing training-congruent effects on a new set of materials. Self-report and interview measures were also used to assess index of bias and state-anxiety before and after training. Importantly, this study shows that experimental modification of attentional bias to threatening information reduces anxiety using laboratory methods independent of a therapist.

Another similarity between attentional bias and the memory biases we discussed in Section 4.2 above is the performance of 'normal', non-anxious controls. As before, it seems that most of us have an adaptive or protective bias in the opposite direction to that of emotionally disordered patients. Look at Figure 16.15 again. Controls are faster in neutral areas than in threat areas, and this has also been found in several subsequent studies. It may be that this represents active avoidance of minor, insignificant threats, such as words and pictures. It would clearly be adaptive to avoid the many distractions of minor threats and single out only serious threats for particular attention.

Now try Activity 16.7.

Although we have mostly mentioned anxious patients so far, biases favouring attention to threat can also be found in high trait anxious participants, although less reliably. Moreover, these biases tend to be stronger when high state anxious mood and high trait anxiety occur together. Anxious patients tend to have relatively high levels of state anxiety much of the time so it is unsurprising that attentional biases are more robust in this group.

4.4 Semantic interpretation

Semantic interpretation (see Chapter 6) is another cognitive process known to be influenced by emotion.

✱ ACTIVITY 16.7

Can you think of a situation where you paid attention to negative cues in your environment because of a fear that you have? If you haven't noticed this negative bias in yourself, have you ever noticed another, non-negative, attentional bias towards features in the environment?

COMMENT

Anyone who is very afraid of spiders might recognize this characteristic in themselves. Almost invariably they will notice any spider in the surroundings well before their non-phobic companions. Most people have a bias to attend to things that match their special interests. Temporary biases are also common, and can, for instance, occur when you have acquired something new such as when you purchase a new car. For a while you may find yourself noticing many examples of this same model, which previously you ignored. This might give you a feel for what it could be like for an anxious individual, although for them it is unpleasant items that just constantly catch their attention without any intention on their part.

If you see a word such as 'batter', do you think of pancakes or do you think of an assault on an innocent victim? It is surprising how many situations in life can be ambiguous and therefore open to biases of interpretation. In this section we shall consider interpretation of ambiguous linguistic information, but be aware that the same processes apply in many situations. Assessing the nature of the shadow in the path ahead on a dark night, or guessing the meaning of the probing look of the interviewer when you apply for a job, are just two other such examples.

The earliest work on interpretation and emotion used **homophones**. These are words like 'pane' and 'pain' or 'die' and 'dye', which sound the same but have different spellings associated with different meanings. Eysenck *et al.* (1987) asked both high and low trait-anxious individuals to write down the homophones as they heard them. All the homophones had both a threatening/unpleasant and a non-threatening or neutral meaning. This simple technique revealed which interpretation had been made by the spelling the participants chose. Eysenck *et al.* found that the higher the participant's trait anxiety, the more threat spellings they produced. This indicated that trait anxiety was linked to a tendency to assume the negative interpretation of an ambiguous stimulus – i.e. an *interpretative bias*.

However, this method soon fell foul of criticism. For example, it is possible that participants were aware of and had access to both spellings, but just chose to write down the negative one. This matters because, if true, it would mean that there was no bias in the actual *interpretation* of the words – both interpretations were made. Instead the bias would be at the stage of making the response, which then says little about the cognitive processing involved in making interpretations. This tendency to report (and not truly interpret) emotional information is called **response bias**.

Later work used an alternative method to avoid this and other problems. For example, in their classic study Richards and French (1992) used **homographs** instead of homophones. These are words that have dual meanings, despite having the same spelling, such as 'batter', 'punch', and 'stalk'. They used these words in a priming experiment involving a lexical decision task (a task described in Chapter 2, Section 1.3). This task involves simply identifying, as rapidly as possible, whether the second of two sequentially presented items is a real word or a meaningless letter string (a non-word). From the participant's point of view the first item that appears is just to be ignored. However, this first word is actually a prime. As described in Chapter 2, if the prime is related in meaning to the second word (the target), as in cat–dog or nurse–doctor, then lexical decisions are expected to be speeded compared to primes and targets that bear no semantic relation (e.g. cat–doctor, nurse–dog).

We can use this logic to infer how participants interpreted the homograph primes. For example, if lexical decisions for trials like batter–assault were faster than for trials like batter–pancake, this would imply that the participant interpreted batter as 'assault' rather than 'pancake'. The results of the Richards and French study, as well as other similar studies, suggest that high anxious participants show a negative bias in interpretation – that is, there is a greater priming effect for target words related to the negative meaning of the homograph than the neutral meaning. For non-anxious participants there is, once again, the familiar positive bias towards the more positive or non-threatening meaning. Further studies have extended this research by using ambiguous sentences or even passages of text, for example:

> *'The doctor examined little Emily's growth' (her height or her tumour?)*
>
> *'The two men watched as the chest was opened' (a gruesome operation or an exciting find?)*
>
> *'Your friend asks you to give a speech at her wedding reception. You prepare some remarks and when the time comes you get to your feet. As you speak, you notice some people in the audience start to laugh' (appreciatively or rudely?)*

Two types of technique used successfully to date are homographs and ambiguous passages of text. Grey and Mathews (2000) repeatedly trained volunteers using a cognitive bias modification of interpretations programme (CBM-I) to interpret the threatening meanings of homographs. They found faster completions of word fragments and lexical decisions when stimuli reflected threatening meanings. Mathews and Mackintosh (2000) used ambiguous passages of text to train participants in either a positive or negative direction. For example, the ambiguous scenario '… as you finish your presentation your employer looks up …' would be induced to be either positive (… approvingly) or negative (… disapprovingly). Participants later presented with novel ambiguous information are then found to spontaneously interpret new material consistent with induction direction. Importantly, it has been reported that interpretative biases are spontaneously deployed at least 24 hours later (Yiend *et al.*, 2005); that induction training survives environmental and contextual changes (Mackintosh *et al.*, 2006); and that positive interpretation induction produces lasting reductions in both trait anxiety levels (Mathews *et al.*, 2007) and vulnerability to an external stressor (Hoppitt *et al.*, 2010).

The concept of protective processing styles such as these has been described formally in a theory known as **attribution theory**. A common observation is that we attribute good things internally, as something within our control, whereas bad things are attributed externally to others or to circumstances. This reflects a tendency to accept the credit for good outcomes and blame something or someone else for bad outcomes. For example, if you are late for an important meeting or fail your driving test you might say 'I'm terribly sorry but the train times have changed and I couldn't help being late' or 'I had such an unreasonable examiner' or 'My instructor gave me inadequate preparation'. If you are early or pass your test first time, you might well congratulate yourself on your efficient organization and planning, or excellent driving skills. You may have come across this described elsewhere as the **self-serving attribution bias** (see Activity 16.8).

Although these self-serving biases might seem an irrational way of thinking, the evidence repeatedly supports their existence and, as with other positive biases, they may have protective properties. Moreover, in emotional disorders, particularly in depression or anxiety, we know that this self-serving bias can be lost or even reversed. Such people might think passing the driving test was just luck, or the examiner being lenient, whereas failing was yet more evidence of their own worthlessness and lack of skill. In some situations it can be shown that by lacking the positive bias, the depressed person's attribution of their own performance can be more accurate than for non-depressed controls, so-called 'depressive realism'.

It should be noted that, although the various positive biases that we have described are thought to be quite normal and have protective qualities (such as helping to maintain good mood and a positive self-image), it is equally true that, taken to their limits, they would be maladaptive.

✱ ACTIVITY 16.8

When you are next chatting with family or friends, or watching conversations on the television, see if you can identify some of the attributions people make. Does this go along with the attribution theory? Do you notice any examples of the self-serving attribution bias?

SUMMARY OF SECTION 4

- State emotion refers to the feelings of the moment whereas traits refer to more enduring personality characteristics of an individual.
- The manifestation of emotion is distinct from the processing of emotional material. The former refers to feelings, behaviours, and bodily responses. The latter refers to the emotional content of the external stimuli upon which the cognitive system acts.

- The field of cognition and emotion is primarily concerned with the conjunction between state or trait emotions and the processing of emotional material.
- Mood congruent memory (MCM) refers to enhanced memory for material that matches present mood. The phenomenon is particularly apparent in depression and may contribute to the clinical disorder.
- Mood dependent memory (MDM) occurs when recall is enhanced by a match between mood at the time of learning and mood at the time of testing. However, the effect is not very robust. Bower's semantic network theory provides one explanation of MDM and MCM.
- Biases in attention and in semantic interpretation are associated with both trait and state anxiety, and again may contribute to chronic anxiety and clinical anxiety disorders.
- Cognitive bias modification procedures experimentally induce biases relevant to pathology-congruent information. Beneficial effects of these procedures include reducing maladaptive biases and improving mood.

5 DOES COGNITION INFLUENCE EMOTION?

5.1 A look at some historical answers

Do we laugh *because* we feel happy or is it the laughing itself that *makes* us feel happy? This question has been central to emotion research since its very beginnings back in the 1880s when William James (1843–1910), commonly regarded to be one of the founders of psychology, first considered it. Putting the question another way, is it our experience of the behavioural and bodily responses associated with emotion that make us subjectively feel that emotion? Or do those responses follow on from our subjective experience of emotion?

5.1.1 James–Lange

James's answer to this question in the late nineteenth century was the counterintuitive one. Namely, he argued that we feel fear *because* we run and we experience happiness as a *result* of laughing: the cognitive and experiential side of emotion was a slave to the physiology of emotion. Carl Lange took a very similar position, and so this view became known as the 'James–Lange' theory (see Figure 16.16(a)). Their observation was that behaviour, most especially in a frightening situation, was initiated too rapidly to have arisen from a *feeling* of fear that was subsequently translated into a conscious decision to act. Rather, they felt that behaviour preceded (conscious) cognition, and more precisely that the experience of emotion depended on the behaviour and bodily reaction that followed an event.

More recent studies, such as those of LeDoux (1996) looking at the speed of the startle response to loud noise, indicate that these responses are initiated within a few milliseconds, well before conscious awareness has time to develop. One implication of this way of looking at things is that physiological responses and behaviours must be distinct and occur in unique constellations in order that different emotions actually occur and feel different. Love and fear feel different, according to James, because they result from different physiological signatures.

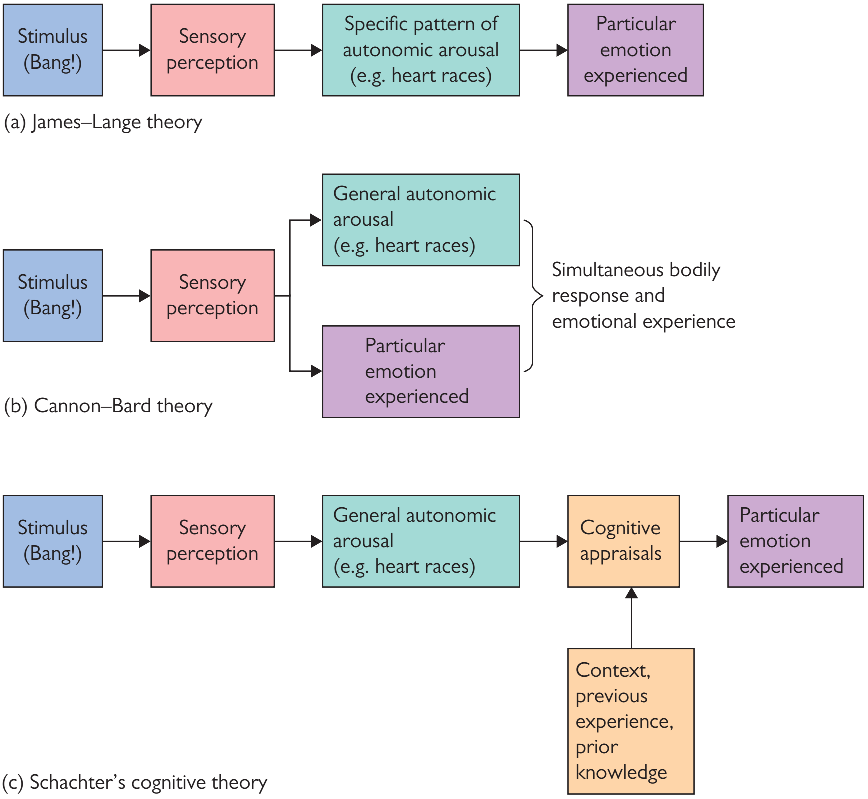

Stimulus (Bang!)
Sensory perception
Specific pattern of autonomic arousal (e.g. heart races)
Particular emotion experienced
(a) James–Lange theory
Stimulus (Bang!)
Sensory perception
General autonomic arousal (e.g. heart races)
Particular emotion experienced
Simultaneous bodily response and emotional experience
(b) Cannon–Bard theory
Stimulus (Bang!)
Sensory perception
General autonomic arousal (e.g. heart races)
Cognitive appraisals
Particular emotion experienced
Context, previous experience, prior knowledge
(c) Schachter's cognitive theory

FIGURE 16.16 Comparing the theories. *Source: Based on Rosenzweig et al., 1999, Figure 15.1, p.412*

5.1.2 Cannon–Bard

Walter Cannon and Philip Bard challenged this view in the 1920s precisely because they felt that physiological responses were pretty indistinguishable across most emotions, and indeed that similar physiological patterns (e.g. increased heart rate, sweating, and inhibited ingestion) could arise from fever during illness. According to them, what distinguishes one emotion from another, given this common physiology, is the pattern of cortical stimulation that arises. For Cannon and Bard, both the autonomic arousal and the subjective experience of a specific emotion could occur simultaneously and were instigated by the higher brain areas such as the cerebral cortex (see Figure 16.16(b)). Thus for Cannon–Bard you don't have to cry (or be suppressing the tendency to do so) to feel sad – there simply has to be an appropriate activation of the thalamus. For the James–Lange theory, preventing crying (and any urge to cry) would prevent sadness.

Both sets of theorists could provide evidence to support their view. The Cannon–Bard camp challenged the view that physiological responses were sufficiently unique to distinguish between the emotions (or indeed between emotion and other causes). They also reasoned that animals and humans with damaged spinal cords, preventing normal physiological responses, nevertheless responded emotionally. In support of the James–Lange theory, emotional response does seem to be blunted in those unfortunate enough to have suffered spinal injury preventing both physiological changes and overt behaviour in response to emotional situations, and this occurs in proportion to the loss of sensation. Emotional feeling is not, however, absent. There must, therefore, be more to emotional feeling than interpreting the sensation of body movement and physiological responses. The debate about the specificity of physiological responses to each emotion, however, still continues. Ekman and colleagues have concluded that there are emotion-specific physiologies for at least anger, fear, disgust, and perhaps sadness (Ekman *et al.*, 1983; Levenson *et al.*, 2001). For example, Ekman claims that skin conductance (discussed in Box 16.1) is higher during sadness than the other emotions, and heart rate decelerates during disgust.

However, John Cacioppo, a renowned psychophysiologist, disagrees. Having reviewed all the available data he concluded (Cacioppo *et al.*, 2000) that the evidence for emotion-specific physiology is far more uncertain and suggests that discrete emotions cannot be differentiated by autonomic activity alone. However, he does agree that there may be a consistent distinction between the positive and negative emotions in general. He proposes that negative emotions are associated with greater motivational output than positive emotions and therefore show generally greater levels of autonomic activation.

5.1.3 Schachter–Singer

When cognitive psychology began to take hold in the 1960s, Stanley Schachter and Jerome Singer proposed an alternative to both the earlier James–Lange and Cannon–Bard views. Like James they held that physiological mechanisms were crucial, but like Cannon they also believed that these responses were non-specific and could not distinguish the different emotions. Instead they thought differentiation was achieved by the individual's particular interpretations or attributions about why those bodily responses were occurring. These ambiguous messages from the body were interpreted by taking into account things like the situational context, previous experience of when certain emotions occur, expectations, and intellectual knowledge of the world. Physiological arousal may be responsible for feeling emotion but cognitive interpretations, or **cognitive appraisals**, were what distinguished one emotion from another (see Figure 16.16(c)).

The Schachter–Singer theory predicts that it should be possible to change our experience of emotion by changing the cognitive appraisals we make, even if the physiological signs remain the same. To test this they performed a now famous experiment (Schachter and Singer, 1962). Participants were injected with epinephrine (adrenalin), which is a hormone that stimulates activity in the sympathetic nervous system (discussed in Section 1.1.2). Some participants were told that there would be no effect of this injection, while others were told that it would make their heart race. The latter group did not report any emotional experience, while the former group did. What does this result say about the James–Lange theory of emotion? The fact that those expecting no effect of injection did actually experience emotion is consistent with James–Lange – the physiology directly led to an emotional experience. However, the other result is inconsistent with that theory. Those able to attribute the bodily sensations to the injection failed to experience emotion. The James–Lange theory makes no allowance for such cognitions to influence emotional experience in this way. It would predict that, despite this knowledge, the physiological arousal should directly give rise to an emotional experience.

In a further aspect of the above experiment, participants were put in a room with a 'stooge' who was party to the experiment and who acted either in an extremely happy or very angry manner. The behaviour of the stooge directly influenced the feelings of those participants who reported experiencing emotion (i.e. those with no foreknowledge of the effects of the injection). Those with the happy stooge reported experiencing happiness, whereas those with the angry stooge reported anger. This demonstrates a strong influence of context on the specific emotion experienced. How does this second finding fit with the James–Lange theory? It is clearly inconsistent with that theory because the same physiological reactions were being experienced differently according to context.

Schachter and Singer used the results of this experiment to support their theory. The non-specific physiological arousal interacted with the social and physical context that participants experienced to determine the precise emotion that was felt. They had succeeded in showing that identical physiological states could be subjectively experienced as different emotions according to how the individual appraised their circumstances. Similarly, subjective emotional experience could be eliminated by telling participants the true source of their bodily sensations. This convincingly showed that cognitive attributions were crucial in whether or not emotions were experienced in their fullest sense.

The Schachter–Singer theory has been criticized. For example, we now know, because we can measure them, that physiological responses associated with different emotions are not in fact identical (and their results have not always been replicated; e.g. Reisenzein, 1983). However, the lasting contribution of this theory was the notion of cognitive appraisal being critically involved in the generation of emotion. The acceptance of this possibility spawned a whole generation of appraisal theorists.

5.1.4 Appraisal theories today

The central idea of appraisal theories is that emotions are elicited and differentiated on the basis of the individual's subjective evaluation of the external situation or event combined with their own physiological state. Although this answers the question of how we differentiate between emotions, it simultaneously raises others. How are these appraisals made – using what yardstick, measuring tool, or criteria?

Each different appraisal theory tends to suggest different dimensions that we use when making appraisals. For example, Klaus Scherer (1999) proposes specific fixed sets of criteria that we supposedly apply to any situation that comes our way. For example, one criterion is the intrinsic qualities of the event, such as how novel or agreeable it is. Another is how significant the event is in terms of our own personal goals or needs. Clearly an event that is neutral in terms of our goals or needs is unlikely to generate emotion. To get an idea of how his criteria can distinguish one emotion from another, look at Table 16.3.

TABLE 16.3 Scherer's appraisal criteria and profiles for different emotions

Stimulus evaluation checks	**Anger/ rage**	**Fear/ panic**	**Sadness**
Novelty			
• Suddenness	High	High	Low
• Familiarity	Low	Various	Low
• Predictability	Low	Low	Various
Intrinsic pleasantness	Various	Various	Various
Goal significance			
• Concern relevance	Order	Body	Various
• Outcome probability	Very high	High	Very high
• Expectation	Dissonant	Dissonant	Various
• Conduciveness	Obstruct	Obstruct	Obstruct
• Urgency	High	Very high	Low
Coping potential			
• Cause: agent	Other	Other/ nature	Various
• Cause: motive	Intent	Various	Chance/ neg.
• Control	High	Various	Very low
• Power	High	Very low	Very low
• Adjustment	High	Low	Medium
Comparability with standards			
• External	Low	Various	Various
• Internal	Low	Various	Various

Note: 'Various' = different appraisal results are compatible with the respective emotion.
Source: Dalgleish and Power, 1999, Table 30.2, p.639

For example, in the table, items under the fear/panic category suggest that this emotion results from an event (stimulus) that is judged to be 'high' on novelty/suddenness and 'low' on novelty/predictability. It is of concern to the body's status (goal significance/concern relevance), urgency is very high, coping potential/power is very low, and so on. For a number of the appraisals involving fear/panic there can be various options; for instance, under novelty/familiarity it is possible to be afraid of something either familiar or unfamiliar, so that this is not a defining feature for that emotion. To give a concrete example, imagine the consequence of a spider emerging from under the sofa for a spider phobic. This is a sudden event, but spiders are familiar (although disliked); they behave unpredictably; their intrinsic pleasantness is 'various' (that is, some people find them pleasant, although a spider phobic certainly would not); the ability to cope is perceived as low, and so on. You may feel that it would be impossible to evaluate an event on all these various things *before* feeling the emotion; there surely would not be enough time? This is a fair criticism and one that has been made by opponents of appraisal theory. However, the counter-argument to this is that appraisals do not have to be conscious serial processes; they may well occur in parallel, automatically.

The evidence in support of appraisal theory relies entirely on subjective self-report and for this reason these theories have been heavily criticized. Typically, participants are either asked to remember personal events, or are exposed to experimental manipulations designed to induce an emotion. They are then asked to report, either verbally or using questionnaires, the types of appraisals they engaged in. You may have spotted another problem with this. By definition, participants will not be able to report on any appraisals made unconsciously, and this is a second major criticism of the evidence for appraisal theory. As yet, appraisal theorists have only been able to counter this by stating that no alternative to self-report exists. It will be interesting to see whether appraisal theorists will be able to find ways of accessing automatic evaluations (perhaps using brain-imaging technologies).

5.2 A clash of minds: the cognition/emotion debate

No discussion of cognition and emotion would be complete without considering one famous example of the different approaches psychologists can take. The two main protagonists in the debate were Richard Lazarus and Robert Zajonc (pronounced zy-unce, to rhyme with once).

5.2.1 Zajonc's view

Zajonc disagreed with appraisal theory's contention that emotions are produced by cognitive processes. He challenged the appraisal theorists directly (Zajonc, 1980) making two key assertions:

1. Appraisal is not necessary for emotion to be experienced. Emotions could arise directly without the need for cognitions at all. (This is similar to the James–Lange idea in that cognition plays no part in the process of eliciting emotion.)
2. The experience of emotion always precedes one's cognitive processing of that emotion. This stronger claim adds to the first by saying that not only is appraisal not necessary, in fact it never occurs before the emotional experience itself. This question of whether emotion precedes cognition, or the other way round, is known as the **primacy debate**.

The following quotation summarizes Zajonc's position very well: 'Affect [meaning mood or state emotion] and cognition are separate and partially independent systems and … although they ordinarily function conjointly, affect could be generated without a prior cognitive process' (Zajonc, 1984, p.117).

How is this issue of primacy different from William James's question? Do we laugh because we're happy or are we happy because we laugh? James was concerned with the relationship between the conscious feeling of experiencing an emotion and the physiological and behavioural expression of that emotion (see Sections 1.1.1 and 1.1.2). The concept of cognitive appraisal had not yet been articulated. The primacy debate contrasts the cognitive appraisal of an emotion with all its other aspects (feeling, physiology, and behaviour).

In support of his argument Zajonc described an experiment that used the famous **mere exposure** effect. Mere exposure refers to the finding that people tend to prefer items to which they have previously been exposed over comparable novel ones. Simple familiarity with something creates a preference for that item. This is presumably one reason for the success of the advertising industry. Zajonc took the mere exposure method and adapted it so that items were presented **subliminally** (below the level of conscious awareness) while participants were engaged in another, primary task. His results revealed that while participants showed no recognition of the subliminal items, they nevertheless gave them higher preference ratings than novel items!

Zajonc argued that these results showed that cognition was not necessary in order to have affective experience. He was assuming, first, that stimuli were not being processed 'cognitively' because they were presented subliminally. Second, he was assuming that preference ratings were tantamount to emotional experience. Both these assumptions have since been challenged. Today the details of non-conscious processing (outside awareness) are controversial, but few would challenge its existence (see Chapter 18). Certainly, it is unlikely that many psychologists now accept Zajonc's implicit assumption that all cognitive processing must be conscious. Likewise, equating preference judgements with affect or emotion is probably a step too far. Surely only very limited emotion is involved in rating how much you like something that has no particular meaning or relevance to you?

5.2.2 Lazarus's view

Richard Lazarus, on the other hand, argued that cognitive appraisal was essential for the experience of emotion: 'Cognitive appraisal (of meaning or significance) underlies and is an integral feature of all emotional states' (Lazarus, 1982, p.1021). In support of his position he undertook several studies. Typically, emotions would be elicited by showing participants anxiety-provoking films. For example, one was a Stone Age circumcision ritual; another showed someone involved in a gruesome industrial accident (it is unlikely that this type of material would obtain ethical approval for use today!). Cognitive appraisal was manipulated by playing one of two soundtracks while participants watched the films. A 'denial' soundtrack included statements indicating that one was a safety film, the people in the film were actors, and the ritual in the film was not actually painful. An 'intellectualization' soundtrack emphasized an anthropological perspective and advocated, for example, considering the ritual as a strange native custom. A control condition had no soundtrack. Physiological measures such as GSR (galvanic skin response) and heart rate were taken throughout viewing and suggested that the appraisals produced by the soundtracks did indeed reduce emotional responses significantly compared with the control condition. Although impressive, these results did *not* prove that cognition necessarily precedes affect, but rather that cognitive appraisal *can* convincingly alter emotional response.

5.2.3 A resolution?

Despite the ferocity of their debate about primacy, neither protagonist marshalled sufficient evidence to win the argument. Rather, a resolution was reached by both identifying their positions more clearly. Zajonc acknowledged the view that the existence of non-conscious appraisal was a key question, and Lazarus conceded that although appraisal might influence emotion this did not mean it was an essential component. Both agreed that, as Zajonc puts it: 'It is a critical question for cognitive theory and for theories of emotion to determine just what is the minimal information process that is required for emotion. Can untransformed pure sensory input directly generate emotional reactions?' (Zajonc, 1984).

Interestingly, further work by Joseph LeDoux (LeDoux, 1989; LeDoux, 1996) has thrown further light on the issue of primacy, suggesting that Zajonc may be right after all. These studies used **lesioned** animals in which specific neural pathways within the animals' brains were deliberately severed by the experimenter. Doing this allows an experimenter to deduce the function of the damaged pathways or regions by giving the animal various tasks to perform and establishing which of these are impaired. You may wish to think about the ethical issues such procedures raise, though we do not have room to consider them here.

Using a variety of tasks manipulating emotions, especially fear, LeDoux has shown that certain brain structures such as the thalamus and the amygdala play different roles in the generation of emotion (see Figure 16.17). Anatomical work has shown that these areas

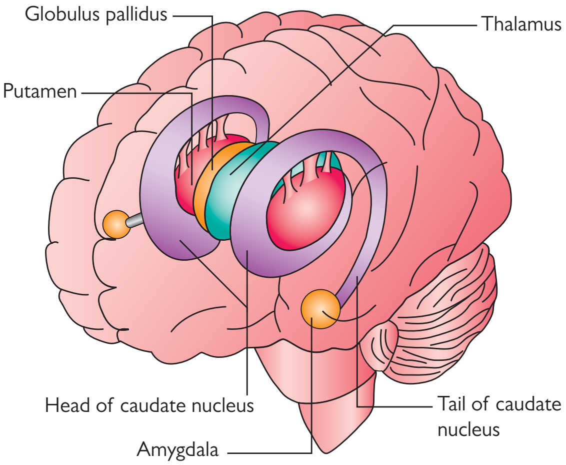


FIGURE 16.17 A schematic drawing of the human brain showing thalamus and amygdala. *Source: Calder et al., 2001, Figure 1(a), p.353*

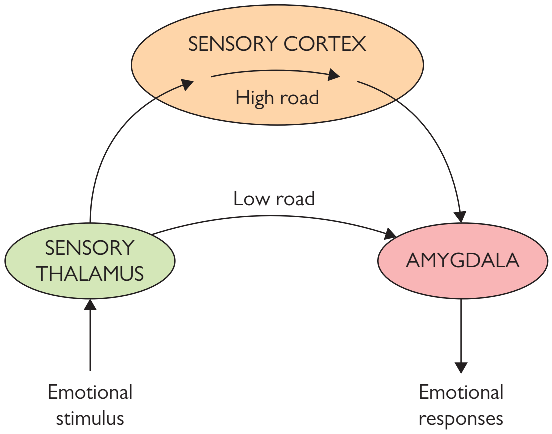


FIGURE 16.18 The low road and the high road to the amygdala. *Source: LeDoux, 1996, Figure 6.13, p.164*

are connected via two routes, as you will see from Figure 16.18. The 'lower' route – so-called because only *evolutionarily old* structures are involved – takes sensory information from the primary sensory areas (the regions of the brain where sensory information arrives first) to the thalamus and then directly to the amygdala. This route bypasses the higher brain structures in the cortex and provides a fast thalamo-amygdala connection involving only one **synapse** (a relay junction between one nerve cell and the next). The 'higher' route – so-called because the *evolutionarily newer* areas such as the cortex are involved – relays information through a more complex route from the thalamus via the sensory cortex to the amygdala.

LeDoux has shown that learning about new fearful situations or altering existing knowledge about fear requires the higher thalamo-cortical-amygdala route to be intact and functioning. However, once fear responses are well learned, lesions to this higher route do not diminish the response. In such cases the lower thalamo-amygdala route is sufficient. He suggests that the lower route is 'quick and dirty', providing a means of rapid identification of, and initiation of responses to, emotionally significant stimuli without the need for time-consuming higher processing. The higher route, on the other hand, is needed to learn new associations or re-learn or extinguish old ones. It is vital in the initial stages of learning about novel fear stimuli, or in the modification of original learning experiences, but when this learning has become automated, the lower, more rapid route seems capable of taking over.

You can see then that this quick and dirty route could be a neuroanatomical substrate (physical basis) for Zajonc's idea of the direct elicitation of emotion without the need for cognition. Maybe the primacy of emotion over cognition was partially right after all. However, the role of the higher route also seems to map onto Lazarus's idea that cognitive processing can precede, or at least influence, emotion. LeDoux suggests that the higher cortical route is necessary to override the quick and dirty route in certain situations – perhaps where the threat turns out not to be so bad, once cognitively processed more fully, or where past knowledge and experience suggest that an emotional reaction would be inappropriate. As is often the case in these debates, both the positions of Lazarus and Zajonc may turn out to be partially correct.

5.2.4 Emotion regulation

Emotion regulation, the process by which we manage and redirect our emotions, is a vital component of everyday human functioning. Emotion regulation involves the active increasing, decreasing, or maintaining of one's emotional state and it is therefore not surprising that vast individual differences have been observed across the lifespan (Koole, 2010). Developmentally, emotion regulation is first influenced by others (such as parents) but is later governed by self-regulatory strategies that can either be automatic or controlled. Gross and Thompson (2007) have identified five common strategies including avoiding unpleasant situations ('situation selection'), adapting situations to make them more tolerable ('situation modification'), and altering attention ('attentional deployment'). Of recent interest, **up-regulation** and **down-regulation** processes refer to an individual's ability to *increase* or *reduce* an emotional response, respectively. In everyday life, up-regulation most often applies to positive emotions (e.g. making an effort to feel happy and relaxed when out partying with friends) whereas down-regulation frequently applies to negative emotion (e.g. suppressing a bad or unhelpful feeling). In research the reverse effects are also of interest, for example up-regulation of neuroticism in high-worriers (Gross, 2007). Emotion regulation can be measured either by self-report or physiological response. Emotion regulation could be a key aspect of adaptive functioning in which most individuals seek to up-regulate pleasure-related emotional responses.

SUMMARY OF SECTION 5

- A question that has always concerned psychologists is whether emotions arise before or after the cognitive processing of the stimuli and situation that elicits them. This is sometimes known as the primacy debate.
- Zajonc argued that emotions do not require prior cognition, whereas Lazarus maintained that cognitive appraisal was an integral part of the production of emotion.
- Recent neuroanatomical and lesion studies suggest that both may have been right after all. LeDoux argues for a 'quick and dirty' neural pathway enabling immediate response to potentially emotional stimuli, without the need for prior cognitive processing. He also proposes a slower cortical route enabling learning and modulation of emotional responses according to the outcome of cognitive appraisals.
- Emotion regulation is the process by which we manage our emotions using self-regulatory strategies. These strategies differ between individuals and can be assessed using self-report and physiological measures.

6 GENERAL SUMMARY

In this chapter we began by noting that emotions comprise at least three components: feelings that can only be reported through introspection; behaviours that can be observed; and bodily responses, some of which can be precisely measured using psychophysiological techniques. We discussed two different approaches to classifying emotions: that of dividing emotional experience up into separate 'basic' emotions, and that of using two or more dimensions to describe a continuum of experience. We briefly explored what some of the functions of emotions might be, such as their use for survival, to enhance performance, or as information signals to tell us how to behave (the somatic marker hypothesis).

For the rest of the chapter we considered the interaction between emotions and cognition, first by looking at how emotional material is processed differently from non-emotional material. In memory, attention, and semantic interpretation we saw how biases in processing usually operate to favour the processing of emotionally significant information. This is a particularly important topic because psychologists believe such biases contribute to emotional disorders such as anxiety and depression. Second, we saw how cognitive appraisals can influence the experience of emotion, and several theoretical variations on this basic idea were described.

Emotion and its interaction with cognition is becoming an increasingly popular area of psychology, helped in part by the availability of new brain-imaging technologies. Armed with the knowledge you have gleaned from this chapter you will be in an excellent position to follow the progress of this exciting area of psychology.

FURTHER READING

Blanchette, I. and Richards, A. (2010) 'The influence of affect on higher level cognition: a review of research on interpretation, judgement, decision making and reasoning', *Cognition and Emotion*, vol.24, no.4, pp.561–95. A critical examination of individual differences in higher level cognitive processes.

Dalgleish, T. and Power, M.J. (1999) *Handbook of Cognition and Emotion*, Chichester, Wiley. A comprehensive review of theories of emotion and their clinical application. Systematically covers each of the Big Five emotions.

Fox, E. (2008) *Emotion Science*, New York, Palgrave Macmillan. An interdisciplinary approach to emotions and emotional disorders that covers both cognitive and neuroscientific techniques.

LeDoux, J.E. (1996) *The Emotional Brain*, New York, Simon and Schuster. A succinct and readable summary of the author's theory, illustrating the integration of bottom-up and top-down approaches to explaining the effects of fear.

Williams, J.M.G., Watts, F.N., MacLeod, C., and Mathews, A. (1997) *Cognitive Psychology and Emotional Disorders*, 2nd edition, Chichester, Wiley. The classic text on cognitive biases and their significance for clinical disorders.

REFERENCES

Amir, N., Beard, C., Burns, M., and Bomyea, J. (2009) 'Attention modification program in individuals with generalized anxiety disorder', *Journal of Abnormal Psychology*, vol.118, no.1, pp.28–33.

Barrett, L. (2006) 'Solving the emotion paradox: categorization and the experience of emotion', *Personality and Social Psychology Review*, vol.10, no.1, pp.20–46.

Bower, G.H. (1981) 'Mood and memory', *American Psychologist*, vol.36, pp.129–48.

Bradley, M.M., Cuthbert, B.N., and Lang, P.J. (1999) 'Affect and the startle reflex', in Dawson, M. E., Schell, A., and Boehmelt, A. (eds.) *Startle Modification: Implications for Neuroscience, Cognitive Science and Clinical Science*, New York, Cambridge University Press.

Cacioppo, J.T., Bernston, G.G., Larsen, J.T., Poehlmann, K.M., and Ito, T.A. (2000) 'The psychophysiology of emotion', in Lewis, M. and Haviland-Jones, J.M. (eds.) *Handbook of Emotions*, 2nd edition, London, The Guilford Press.

Calder, A.J., Keane, J., Manes, F., Antoun, N., and Young, A.W. (2000) 'Impaired recognition and experience of disgust following brain injury', *Nature Neuroscience*, vol.3, pp.1077–8.

Calder, A.J., Lawrence, A.D., and Young, A.W. (2001) 'Neuropsychology of fear and loathing', *Nature Reviews Neuroscience*, vol.2, no.5, pp.352–63.

Dalgleish, T. and Power, M.J. (1999) *Handbook of Cognition and Emotion*, Chichester, Wiley.

Damasio, A.R. (1996) 'The somatic marker hypothesis and the possible functions of the prefrontal cortex', *Philosophical Transactions of the Royal Society of London, Series B*, vol.351, pp.1413–20.

Darwin, C. (1998, first published 1872) *The Expression of the Emotions in Man and Animals*, 3rd edition, London, Harper Collins.

Dawson, M.E., Schell, A.M., and Bohmelt, A.H. (eds.) (1999) *Startle Modification*, Cambridge, Cambridge University Press.

Eibl Eibesfeldt, I. (1988) 'Social interactions in an ethological, cross-cultural perspective', in Poyatos, F. (ed.) *Cross Cultural Perspectives in Nonverbal Communication*, Kirkland, WA, Hogrefe & Huber Publishers.

Eich, E. and Metcalfe, J. (1989) 'Mood dependent memory for internal versus external events', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.15, pp.443–55.

Ekman, P. (1973) 'Universal facial expressions in emotion', *Studia Psychologica*, vol.15, pp.140–7.

Ekman, P. (1992) 'An argument for basic emotions', *Cognition and Emotion*, vol.6, pp.169–200.

Ekman, P. (1999) 'Facial expressions', in Dalgleish, T. and Power, M.J. (eds.) *Handbook of Cognition and Emotion*, New York, John Wiley & Sons Ltd.

Ekman, P. and Friesen, W.V. (2003) *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*, Cambridge, MA, Malor Books.

Ekman, P. and Oster, H. (1979) 'Facial expressions of emotion', *Annual Review of Psychology*, vol.30, pp.527–54.

Ekman, P., Levenson, R.W., and Friesen, W. (1983) 'Autonomic nervous system activity distinguishes among emotions', *Science*, vol.221, pp.1208–10.

Ekman, P., Sorenson, E.R., and Friesen, W.V. (1969) 'Pan-cultural elements in facial displays of emotion', *Science*, vol.164, pp.86–8.

Eysenck, M. and Keane, M.T. (1996) *Cognitive Psychology: A Student's Handbook*, 3rd edition, Hove, Psychology Press.

Eysenck, M.W., MacLeod, C., and Mathews, A.M. (1987) 'Cognitive functioning in anxiety', *Psychological Research*, vol.49, pp.189–95.

Frijda, N.H. (2001) 'The self and emotions', in Bosma, H.A. (ed.) *Identity and Emotion: Development Through Self Organization, Studies in Emotion and Social Interaction*, New York, Cambridge University Press.

Frijda, N.H. and Tcherkassof, A. (1997) 'Facial expressions as modes of action readiness', in Russell, J.A. and Fernández Dols, J.M. (eds.) *The Psychology of Facial Expression*, *Studies in Emotion and Social Interaction*, *Second Series*, New York, Cambridge University Press.

Grey, S. and Mathews, A. (2000) 'Effects of training on interpretation of emotional ambiguity', *Quarterly Journal of Experimental Psychology*, vol.53, no.4, pp.1143–62.

Gross, J.J. (2007) *Handbook of Emotion Regulation*. New York, Guilford Press.

Gross, J.J. and Thompson, R.A. (2007) 'Emotion regulation: conceptual foundations', in Gross, J.J. (ed.) *Handbook of Emotion Regulation*, New York, Guilford Press.

Hoppitt, L., Mathews, A., Yiend, J., and Mackintosh, B. (2010) 'Cognitive mechanisms underlying the emotional effects of bias modification', *Applied Cognitive Psychology*, vol.24, pp.312–25.

Izard, C.E. (2007) 'Basic emotions, natural kinds, emotion schemas, and a new paradigm', *Perspectives in Psychological Science*, vol.2, no.3, pp.260–80.

Joorman, J., Hertel, P., LeMoult, J., and Gotlib, H. (2009) 'Training forgetting of negative material in attention', *Journal of Abnormal Psychology*, vol.118, no.1, pp.34–43.

Koole, S. (2010) 'The psychology of emotion regulation: an integrative review', in De Houwer, J. and Hermans, D. (eds.) *Cognition and Emotion*, East Sussex, Psychology Press.

Lazarus, R.S. (1982) 'Thoughts on the relations between emotion and cognition', *American Psychologist*, vol.37, pp.1019–24.

LeDoux, J.E. (1989) 'Cognitive-emotional interactions in the brain', *Cognition and Emotion*, vol.3, pp.267–89.

LeDoux, J.E. (1996) *The Emotional Brain*, New York, Simon and Schuster.

Levenson, R.W., Cacioppo, J.T., Davidson, R.J., Lang, P., Ohman, A., and Stemmler, G. (2001) 'Psychophysiology of emotion: a decade and a half later', *Psychophysiology*, vol.38, Supplement S4.

Mackintosh, B., Mathews, A., Yiend, J., Ridgeway, V., and Cook, E. (2006) 'Induced biases in emotional interpretation influence stress vulnerability and endure despite changes in context', *Behaviour Therapy*, vol.37, no.3, pp.209–22.

MacLeod, C., Mathews, A., and Tata, P. (1986) 'Attentional bias in emotional disorders', *Journal of Abnormal Psychology*, vol.95, no.1, pp.15–20.

Mathews, A. (1990) 'Why worry? The cognitive function of anxiety', *Behaviour Research and Therapy*, vol.28, no.6, pp.455–68.

Mathews, A. and Mackintosh, B. (2000) 'Induced emotional interpretation bias and anxiety', *Journal of Abnormal Psychology*, vol.109, no.4, pp.602–15.

Mathews, A. and MacLeod, C. (2002) 'Induced processing biases have causal effects on anxiety', *Cognition and Emotion*, vol.16, no.2, pp.331–54.

Mathews, A., Ridgeway, V., Cook, E., and Yiend, J. (2007) 'Inducing a benign interpretational bias reduces trait anxiety', *Journal of Behaviour Therapy and Experimental Psychiatry*, vol.38, no.2, pp.225–36.

Matt, G.E., Vacquez, C., and Campbell, W.K. (1992) 'Mood-congruent recall of affectively toned stimuli: a meta-analytic review', *Clinical Psychological Review*, vol.12, pp.227–55.

Medicus, G., Schleidt, M., and Eibl Eibesfeldt, I. (1994) 'Universelle Zeitkonstante bei Bewegungen taubblinder Kinder' ('Universal time constancy in movements of deaf-blind children'), *Nervenarzt*, vol.65, no.9, pp.598–601.

Oatley, K. and Jenkins, J.M. (1996) *Understanding Emotions*, Oxford, Blackwell Publishers Ltd.

Oatley, K. and Johnson-Laird, P.N. (1987) 'Towards a cognitive theory of emotions', *Cognition and Emotion*, vol.1, pp.29–50.

Ortony, A. and Turner, T.J. (1990) 'What's basic about basic emotions?' *Psychological Review*, vol.97, no.3, pp.315–31.

Panksepp, J. (1989) 'The neurobiology of emotions: of animal brains and human feelings', in Wagner, H. and Manstead, A. (eds.) *Handbook of Social Psychophysiology*, *Wiley Handbooks of Psychophysiology*, Oxford, John Wiley and Sons.

Panksepp, J., Sacks, D.S., Crepeau, L.J., and Abbott, B.B. (1991) 'The psycho- and neurobiology of fear systems in the brain', in Denny, M.R. (ed.) *Fear, Avoidance, and Phobias: A Fundamental Analysis*, Hillsdale, NJ, Lawrence Erlbaum Associates Inc.

Plutchik, R. (1991) *The Emotions*, New York, University Press of America.

Plutchik, R. and Landau, H. (1973) 'Perceived dominance and emotional states in small groups', *Psychotherapy: Theory, Research and Practice*, vol.10, pp.341–2.

Power, M.J. and Dalgleish, T. (1997) *Cognition and Emotion: From Order to Disorder*, Hove, Psychology Press.

Reber, A.S. (1995) *The Penguin Dictionary of Psychology*, 2nd edition, Harmondsworth, Penguin.

Reisenzein, R. (1983) 'The Schachter theory of emotions: two decades later', *Psychological Bulletin*, vol.94, pp.239–64.

Richards, A. and French, C.C. (1992) 'An anxiety-related bias in semantic activation when processing threat/neutral homographs', *Quarterly Journal of Experimental Psychology*, vol.45, pp.503–25.

Rosenzweig, M.R., Leiman, A.L., and Breedlove, S.M. (1999) *Biological Psychology: An Introduction to Behavioural, Cognitive and Clinical Neuroscience*, 2nd edition, Sunderland, MA, Sinauer Associates Inc.

Schachter, S. and Singer, J. (1962) 'Cognitive, social and physiological determinants of emotional state', *Psychological Review*, vol.69, pp.379–99.

Scherer, K.R. (1999) 'Appraisal theories', in Dalgleish, T. and Power, M. (eds.) *Handbook of Cognition and Emotion*, Chichester, Wiley.

Scherer, K.R. and Wallbott, H.G. (1994a) 'Evidence for universality and cultural variation of differential emotion response patterning', *Journal of Personality and Social Psychology*, vol.66, no.2, pp.310–28.

Scherer, K.R. and Wallbott, H.G. (1994b) 'Evidence for universality and cultural variation of differential emotion response patterning: correction', *Journal of Personality and Social Psychology*, vol.67, no.1, pp.55.

Scherer, K.R., Banse, R., and Wallbott, H. (2001) 'Emotion inferences from vocal expression correlate across languages and cultures', *Journal of Cross-Cultural Psychology*, vol.32, no.1, pp.76–92.

Segal, Z.V., Williams, J.M.G., and Teasdale, J.D. (2002) *Mindfulness-Based Cognitive Therapy for Depression: A New Approach to Preventing Relapse*, New York, Guilford Press.

Stroop, J.R. (1935) 'Studies of interference in serial verbal reactions', *Journal of Experimental Psychology*, vol.18, pp.643–62.

Teasdale, J.D. (1988) 'Cognitive vulnerability to persistent depression', *Cognition and Emotion*, vol.2, no.3, pp.247–74.

Wallbott, H.G. and Scherer, K.R. (1988, first published 1986) 'How universal and specific is emotional experience? Evidence from 27 countries on five continents', in Scherer, K.R. (ed.) *Facets of Emotion: Recent Research*, Hillsdale, NJ, Lawrence Erlbaum Associates Inc. (First published in *Social Science Information* (*Sur les sciences sociales*), vol.25, no.4, pp.763–95.)

Wilson, E., MacLeod, C., Mathews, A., and Rutherford, E. (2006) 'The causal role of interpretative bias in anxiety reactivity', *Journal of Abnormal Psychology*, vol.115, no.1, pp.103–11.

Yerkes, R.M. and Dodson, J.D. (1908) 'The relation of strength of stimulus to rapidity of habit-formation', *Journal of Comparative Neurology and Psychology*, vol.18, pp.459–82.

Yiend, J., Mackintosh, B., and Mathews, A. (2005) 'Enduring consequences of experimentally induced biases in interpretation', *Behaviour Research and Therapy*, vol.43, no.6, pp.779–97.

Zajonc, R.B. (1980) 'Feeling and thinking: preferences need no inferences', *American Psychologist*, vol.35, pp.151–75.

Zajonc, R.B. (1984) 'On the primacy of affect', *American Psychologist*, vol.39, pp.117–23.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

CHAPTER 17

AUTOBIOGRAPHICAL MEMORY AND THE SELF

Clare J. Rathbone, Chris J.A. Moulin, Martin A. Conway, and Emily A. Holmes

1 WHAT ARE AUTOBIOGRAPHICAL MEMORIES?

Consider the following memories:

1 *An account of a personally significant song: 'I can remember being on a bus one day in 1989, returning home from yet another tedious day at work listening to 'Strangeways Here We Come' on my walkman. I got to 'Last Night I Dreamt That Somebody Loved Me' and felt totally in despair as I listened to Morrissey's pitiful and anguished lyrics. I saw my reflection in the bus window, and I immediately wanted to throw myself off and lie down and die, because what I saw frightened me almost to death.'*

(Angie Lewis reported in Gallagher et al., 1995)

2 *A response made by a person asked to recall a memory to the (cue) word 'ship': 'We were going on holiday to France. I remember that we stayed at a boarding house in Dover and went down to the ferry very early the following morning. My brother and I were wildly excited; it was the first time we had been abroad and the first time we had been on a ship of any sort. I have a vivid memory of looking back at the White Cliffs as the boat pulled out of the harbour – they seemed immensely tall.'*

(Conway, 1996)

3 *A memory recalled when reading about 'flashbulb' memories – vivid memories of one's personal circumstances when learning an item of news (Brown and Kulik, 1977): 'My own memory for the declaration of the Second World War, from September 1939, occurred when I was aged 6 years and 6 months. I have a clear image of my father standing on the rockery of the front garden of our house waving a bamboo garden stake like a pendulum in time with the clock chimes heard on the radio which heralded the announcement. More hazily, I have an impression that neighbours were also out in the adjoining gardens listening to the radio and, although my father was fooling around, the feeling of the memory is one of deep foreboding and anxiety.'*

(Gillian Cohen cited in Conway and Pleydell-Pearce, 2000)

4 *A memory reported by David Pillemer in a study of 'self-defining' memories (Singer and Salovey, 1993): 'I remember sitting in X's class on the day that a mid-term . . . was handed back. I was a freshman and felt that I was in over my head. The professor gave a stern lecture on the values of good writing before she handed back the papers. As she reproached us, my terror grew because her comments seemed to be personally directed at me. I was from a small town, did not have the same background as anyone in my class, and had immediately felt my inadequacies when class began in September. Then she said: "But Y has answered the questions well and has an unusual lyrical and personal style that enhanced her answer." I couldn't believe that she was talking about my paper, but she was. I can still envision that dimly-lit little room in the bottom of Z and smell its peculiar musty odour. I can still picture her stern but kind face and feel the relief and pride I felt at that moment.'*

(Pillemer, 1998)

5 *An account of a traumatic experience from a person suffering from post-traumatic stress disorder (PTSD): 'A man who drove cars for a living was involved in a road traffic accident. He was a back seat passenger in a car when it was in a high speed collision with another vehicle; activation of the air bags in the front of the car produced a cloud of powder, which he thought at the time was smoke. At the time he could smell petrol and thought the car might ignite and remembered thinking "I will be burned alive." His wife was unconscious after the impact and he thought that she had died. He remembered thinking to himself "What am I going to do now?" as he thought about his future alone without his wife. He had been experiencing terrific guilt about this as it suggests to him that he is a selfish person. In addition, he was an experienced driver and anticipated the crash, but did not cry out. He felt that he could have averted the crash if he had done this. He experienced intrusive thoughts, such as "I should have shouted" (to warn the driver) and he relived the feeling he felt when he thought his wife had died, which he believed to be his fault because he had not shouted out.'*

(Conway et al., 2004; see also Ehlers et al., 2004)

In autobiographical memories like these, from the mundane to the profound, the involvement of the self is clear. These memories arguably provide a personal historical context or personal biography for who we are now; they are in essence the 'database' of the self (Conway and Pleydell-Pearce, 2000; McAdams, 2001; see also Hollway and Jefferson, 2000). They help us integrate with each other and with the history of our times, and give a continuity to experience that would not otherwise be possible. Such a central form of cognition is, much as one might expect, highly complex and engages processes in many different parts of the brain. Because of this, autobiographical memory is highly susceptible to changes in brain function and is easily disrupted by brain injury, the experience of trauma, and by psychiatric illnesses, all of which are reviewed in this chapter.

The statements above are complex accounts of personal experiences, either freely retrieved or recalled to cues. Such accounts provide insights into commonalities across people and through time – research into autobiographical memory has revealed systematic similarities in people's autobiographical memories, particularly evident in the distribution of memories across the lifespan. You might have already noticed in the example memories listed above that some date to when the rememberers would have been in their late teens or early twenties (for example, memories 1 and 4). This seems to be a time when particularly enduring memories are formed that remain highly accessible, in contrast to memories from childhood and infancy, which are difficult to access. Indeed, in the example memories above none date to when the rememberers were five years or younger. In Section 2 we consider the accessibility of memories across the lifespan. Section 3 concentrates on the psychological nature of autobiographical memories, their representation in long-term memory, and their relation to the self. In Section 4 we review findings on disruptions of autobiographical memory following brain injury and the experience of trauma.

2 AUTOBIOGRAPHICAL MEMORY ACROSS THE LIFESPAN

First of all, try Activity 17.1 overleaf.

Autobiographical memories are complex mental constructions that take time to bring to mind and once in mind have to be effortfully maintained. Although, of course, in abnormal remembering, such as occurs in PTSD (post-traumatic stress disorder; see Section 4), exactly the reverse may occur and some details of a trauma may be spontaneously and intrusively recalled and prove difficult to keep out of mind (such as the experience of guilt in Memory 5 at the beginning of this chapter). Clearly, some set of higher-order executive processes must operate to construct memories appropriately, to keep irrelevant knowledge and memories out of mind, and to ensure that what is recalled is relevant to the task or goal currently active. Such a structure is often called the **working self** (Conway and Pleydell-Pearce, 2000). The choice of name was deliberate and intended to make an explicit connection to the concept of 'working memory' (Baddeley, 1986, 2000; see Chapter 9). The working self is conceived as a hierarchy of currently active goals (**goal hierarchy**) and self-conceptions through which current experience is encoded and in which memories are constructed. Because of this, the self has a profound influence on the accessibility of autobiographical knowledge and therefore upon the process of memory construction. This influence may extend across the lifespan, so that periods of change and development of the self, which contain self-defining memories that are crucial to the self, may be particularly marked in autobiographical memory. The distinguishing aspect of memories and knowledge from these times may lie in their raised accessibility relative to other, more dormant periods; in other words, memories from these periods readily come to mind.

The working self probably first emerges in some more or less coherent form as children develop the ability for objective and subjective self-awareness, i.e. conceptions of 'I' and 'Me', in their second year (Howe and Courage, 1997). Certainly children as young as 30 months have detailed autobiographical memories (Fivush *et al.*, 1996), although these memories tend not to be accessible in adulthood. Arguably, one reason why memories are more accessible at certain times in the lifespan reflects an underlying shift in the self, perhaps only stabilizing into an enduring form in late adolescence and early adulthood (Erikson and Erikson, 1982/1997). These periods of development of the self are reflected in the **lifespan retrieval curve**, which is observed when older adults (about 35 years and older) recall autobiographical memories in free recall or in a variety of cued recall conditions (Franklin and Holding, 1977; Fitzgerald and Lawrence, 1984; Rubin *et al.*, 1986; Rubin *et al.*, 1998). Memories are plotted in terms of age

✱ ACTIVITY 17.1

The lifespan retrieval curve

We are going to carry out an autobiographical memory retrieval experiment. You will need the following equipment: a pen, a stack of plain paper (say 20 to 40 sheets of A4 cut in half), and a watch. You will also need a quiet room to work in for about an hour. When you are ready, read the instructions and start immediately.

Instructions

1. In the next 10 minutes recall as many memories as you can. The memories should be specific and detailed as in the examples at the start of the chapter. Try to sample from across your life and avoid recalling memories all from one period (for example, a recent holiday). No memories from the past 12 months are allowed.
2. Each time you recall a memory, write down a short title on a piece of paper. The title should be designed so that if you read it again you would know exactly what you recalled. IMPORTANT: turn the sheet face down and do not look at the title again during this recall phase.
3. When the 10 minutes are up, STOP.
4. DO NOT READ FURTHER UNTIL YOU HAVE RECALLED YOUR MEMORIES. THEN RETURN TO THE INSTRUCTIONS BELOW.

Now go back through and date each of the memories by recording how old you were, in months, when you think the recalled event took place (Age at Encoding or AaE). If you really want to simulate an autobiographical memory experiment you could also rate each of the memories on the rating scales used in Activity 17.2.

Now we want to plot AaE. To do this, use a scale ranging from '0' (birth) to your actual age now. Then divide the scale into five-year time bins (any size of 'bin' will do and we have chosen five-year 'bins' or periods of time simply because this is often used in published reports). This AaE scale will form the *x*-axis running along the bottom of the graph. The *y*-axis will be a count of the number of memories falling in each five-year time bin and will run from 0 to about 10 (it is unlikely that you will have more than 10 memories falling in any one time bin but if you do, increase the *y*-axis scale to, say, 15 or 20, or whatever best suits your data). For each bin, plot how many memories fall in that time bin. For example, maybe six memories date to the period when you were 20 to 25 years of age. The 'X' marked in this bin will then map on to '6' on the *y*-axis. Next join up the 'X's and compare your lifespan retrieval curve to Figure 17.1.

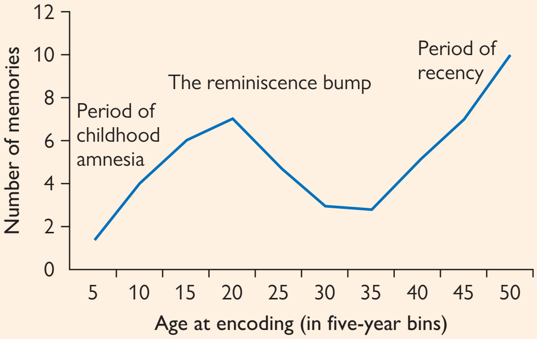


FIGURE 17.1 An idealized representation of the lifespan retrieval curve.

at encoding of the remembered experiences, and the resulting lifespan retrieval curve typically takes a form similar to that shown in Figure 17.1. Did your own lifespan retrieval curve take this form? Experiments in our laboratories repeatedly produce curves such as in Figure 17.1, although we do find that people in their twenties or younger do not show such a clear pattern. As can be seen in Figure 17.1 the lifespan retrieval curve consists of three components: the period of childhood amnesia (from birth to approximately five years of age), the period of the reminiscence bump (from 10 to 30 years) and the period of recency (which declines from the present back to the period of the reminiscence bump).

2.1 Childhood amnesia

There are many theoretical explanations for childhood amnesia (see Pillemer and White, 1989), but most flounder on the fact that children below the age of five years have a wide range of specific and detailed autobiographical memories (Fivush *et al.*, 1996), which are not accessible later in life. Explanations that postulate childhood amnesia to be related to general developmental changes in intellect, language, emotion, etc. fail simply because apparently normal autobiographical memories were in fact accessible when the individual was in the period of childhood amnesia. It seems unlikely that an increase in general functioning would make unavailable already accessible memories.

From the perspective of the Conway and Pleydell-Pearce (2000) model of autobiographical memory this period is seen as reflecting changes in the working self's goal hierarchy. The goals of the infant and young child, through which experience is encoded into memory, are so different from those of the adult that the adult working self is unable to access those memories. Another possibility, one much more in line with Freudian thinking on childhood amnesia (Freud, 1899/1955), is that the working self of infancy and early childhood is much less able to control the occurrence and intensity of emotional experience. Episodic memories encoded during this period are then saturated with intense emotions and, if recalled in maturity, could destabilize the adult working self by reinstating intense infant emotions. This view suggests that access to autobiographical memories encoded during this period might be quite powerfully limited by the adult working self, leading to the lack of memories from this period. Currently, there is no generally accepted explanation for this component of the lifespan retrieval curve, although a recent study has pointed to language factors in the retrieval of early memories. Morrison and Conway (2010) have shown that the AaE of adults' autobiographical memories is related to the age of acquisition (also abbreviated AoA) of the word used to probe that memory. Age of acquisition norms have now been established for many different words (Gilhooly and Logie, 1980; Stadthagen-Gonzalez and Davis, 2006). Morrison and Conway showed that people tended to produce an earlier autobiographical memory in response to words acquired early in life, e.g. jelly, and a later memory for words acquired later in life, e.g. cigarette. This suggests that linguistic factors may play at least some part in the accessibility of memories from earlier in life.

2.2 The reminiscence bump

The second component of the lifespan retrieval curve is the period when rememberers were aged 10 to 30 years, which is known as the reminiscence bump (RB). The RB is distinguished by an increase in recall of memories relative to the period that preceded it and those that follow it. The RB has been observed in dozens of studies, leading David Rubin (a leading researcher in this area) to describe it as one of the most 'reliable' empirical observations in cognitive psychology (e.g. Rubin, 2002). Nonetheless, care must be taken in collecting memories for the RB. If memories are given dates as they are recalled then rememberers have a tendency to become 'stuck' in a time period. Then they may not produce an RB, or may produce an exaggerated RB depending on which time period the rememberer adheres to. Similarly, some rememberers can become 'stuck' in the very recent past and recall only memories from the last few months, again obscuring the RB and period of childhood amnesia. As a result, it is typical to ask people to avoid generating memories for the most recent 12-month period, as in Activity 17.1, to avoid this 'recency effect' biasing retrieval.

In general, these types of retrieval strategies need to be minimized and access to memories should be open (rather than constrained or directed). The rememberer should therefore respond with the first memories that come to mind, i.e. those that are most accessible, for the full lifespan retrieval curve to be observed. When these conditions are met the RB is frequently observed. Interestingly, however, the RB is present not just in the recall of specific autobiographical memories but also emerges in a range of different types of autobiographical knowledge. For example, the RB has been observed in the recall of films (Sehulster, 1996), music (cf. Rubin *et al.*, 1998), books (Larsen, 1998), and public events (Holmes and Conway, 1999; Schuman *et al.*, 1997). Memories recalled from the period of the RB are more accurate (Rubin *et al.*, 1998). They are judged by the rememberer as more important than memories from other time periods, and are rated as highly likely to be included in one's autobiography (Fitzgerald, 1988, 1996; Fromholt and Larsen, 1991, 1992; Rubin and Schulkind, 1997). The autobiographical memories of middle-aged and older adults are therefore characterized by a high degree of accessibility to autobiographical memories dating to the period when they were 10 to 30 years of age, and this is typically most marked for the narrower period of 15 to 25 years of age.

In a rather similar manner to the period of childhood amnesia, the RB also has several plausible explanations (see Rubin *et al.*, 1998). Some obvious explanations can, however, be ruled out. Memories from the RB period are not dominated by first-time experiences, but rather appear to consist of memories of experiences that are idiosyncratic to individual rememberers. Similarly, the suggestion that memories from the RB are more vivid – the idea being that memory encoding is at peak efficiency during this period – turns out to be incorrect (Holmes and Conway, 1999; Janssen *et al.*, 2011; Rubin and Schulkind, 1997). Also incorrect are explanations that either postulate preferential effort to recalling memories from this period, or suggest RB memories are of more pleasant experiences. However, research has shown that the RB is found only for positive (and not negative) events (e.g. Rubin and Berntsen 2003), showing some role of emotion in the RB.

It seems that a more complex explanation is required of the RB, and there have been several different accounts over the years focusing variously on biological, cognitive, and self-based explanations (for a review see Rubin *et al.*, 1998). One recent explanation comes from Rubin and Berntsen and is based upon people generating a 'life script' of their experiences (Rubin and Berntsen, 2003; Berntsen and Rubin, 2004). In their influential 2003 study, Rubin and Berntsen showed that the structure and qualities of people's memories from the RB matched the structure and qualities of 'memories' generated for a hypothetical 70-year-old. This suggests an overlap between the life story structure for someone away from the context of a memory task, and for the memories retrieved of one's own life. You could try emulating this study – have a go at generating a fictitious character and see if you can generate defining events from their life. When do these occur? Rubin and Berntsen (2003) found that, just like 'real' autobiographical memories, they tended to come from the RB, and were emotionally positive. Another theme of Rubin's work is that 'Events from the bump period are remembered best because they occur when rapid change is giving way to relative stability that lasts at least until retrieval' (Rubin, 2002, p.14). By this view, a period of rapid change is dominated by novel experiences that more fully engage encoding processes, and so become represented in memory in a highly accessible way and lead to the RB.

An alternative theory is that the high accessibility of memories from this period may be related to their enduring relation to the self (Conway and Pleydell-Pearce, 2000). Arguably, many memories from the period of the RB are of self-defining experiences (Fitzgerald, 1988; Singer and Salovey, 1993), and have a powerful effect in binding the self to a specific reality. The 'novelty' of RB experiences lies in their distinctiveness and uniqueness *for the self*, and they may play a crucial role in the final formation of a stable self-system during late adolescence and early adulthood. Memories from this period help to define identity (Conway, 1996) and, because of this, they endure in memory in a highly accessible form. Conway's account and Rubin and Berntsen's account are not necessarily in opposition to each other and, as often occurs in psychology, might both be plausible explanations. The life scripts theory might be a good description of the cultural understanding of the lifespan, but it may be that the organizational properties of the self are what drive us to characterize our lives like this. One thing is clear from this research – the reminiscence bump is not purely a memory phenomenon – it occurs when we talk about our favourite books and films and when we put together the life of a fictitious character, and it is even evident when children are asked to imagine the events that might occur during their future life (Bohn and Berntsen, 2011).

2.3 Recency

The final component of the lifespan retrieval curve is the 'recency' component (see Figure 17.1). It is possibly the least interesting and can be simply explained as a period of forgetting older memories: memories recently encoded remain accessible, while memories retained over a longer retention interval are subject to decay and/or interference and so become progressively less accessible. This pattern of retention is familiar from laboratory studies and is one that has been observed many times. On the other hand it might be questioned why such memories or salient experiences should be 'forgotten' in this way. Moreover, it might also be noted that when people are specifically instructed to recall older autobiographical memories, there are apparently plenty of available memories (see Holmes and Conway, 1999, for example). Thus, it is not forgetting but rather a bias or preference in access. It may be that the recency portion of the lifespan retrieval curve reflects a lowering in self-relevance of memory for experiences from the recent past, and hence a corresponding lowering of accessibility rather than complete forgetting. Thus, as recently acquired autobiographical memories become less relevant to the self's goals, their accessibility is attenuated, but not lost, and can be restored by direct attempts to retrieve recent information. Of course, as the retention interval lengthens access may actually become lost, rather than just attenuated, and in that case forgetting would occur.

SUMMARY OF SECTION 2

- The 'lifespan retrieval curve' illustrates how frequently autobiographical memories are recalled over different periods in someone's life. The lifespan retrieval curve is characterized by periods of childhood amnesia, the reminiscence bump, and recency.
- The concept of the 'working self' (Conway and Pleydell-Pearce, 2000) can be thought of as a hierarchy of currently active goals and self-concepts through which experience is encoded and memories constructed.

3 AUTOBIOGRAPHICAL KNOWLEDGE, EPISODIC MEMORY, THE WORKING SELF, AND MEMORY CONSTRUCTION

Try Activity 17.2 before you read any further.

The pattern of memories retrieved over the lifespan has a particular shape, as shown in Figure 17.1, and one that strongly implicates the self in memory retrieval. The lifespan retrieval curve is, however, just one aspect of this complex higher order form of cognition. Another and

✱ ACTIVITY 17.2

Taking part in an autobiographical memory experiment: retrieving memories to cue words

Before reading further, it will be useful to retrieve a few more of your own memories, and reflect on what comes to mind both while forming a memory and when it is fully constructed. To do this, imagine that you are a participant in an autobiographical memory experiment. The experimenter tells you that you will have to bring to mind memories of specific experiences of events that you yourself experienced, and that took place over periods of seconds, minutes, and hours, but no longer than one day – as with the example memories listed at the start of the chapter. This means that responses such as 'last summer', 'when I was little', or 'holiday in Italy' are too general and do not count as memories. Instead you are required to recall detailed memories, memories of specific events. These can be from any part of your life, indeed sampling widely would be good, but they should not be of events experienced in the last 12 months.

You are asked to recall specific memories by reading 'cue' words, then bringing to mind the first memory about which the cue word reminds you – bearing in mind the constraints of sampling widely and not from the past 12 months. Once you have the memory in mind, write down a description of it and provide a title. You should also rate each memory on the following scales:

Let's create a response sheet now. On a sheet of paper write the following:

Memory I

Title: . . ..

Memory description: . . ..

Ratings (circle a number):

Vividness:	1	2	3	4	5
Valence:	1	2	3	4	5
Intensity:	1	2	3	4	5
Rehearsal:	1	2	3	4	5

AaE: . . .. (leave this blank for now)

Do this three times so you have three memory response sheets (in an actual experiment far more memories would be collected, usually 20 or more).

Assuming you are now ready:

1. Recall your first memory to the cue word CHAIR and then complete the response sheet.
2. Now recall a memory to the word ILLNESS and complete the response sheet.
3. Finally recall a memory to the cue SUMMER and complete the response sheet.

TABLE 17.1 Memory vividness

	1	2	3	4	5
Memory vividness	No imagery	Some imagery	Usual image vividness	Vivid imagery	Extremely vivid imagery
Valence of the remembered experience	Very negative	Negative	Neutral	Positive	Very positive
Emotional intensity of the remembered experience	Very mild	Some emotion	Emotional	Intense	As intense as any emotional experience I have ever had
Rehearsal: how frequently have you thought and/or talked about this event?	Very rarely	Sometimes	With about average frequency	Above average	Very frequently

4 Now, go back and at the bottom of each response sheet on the line that says ‘AaE’ write (in months and as exactly as you can) your age when the remembered event occurred (Age at Encoding or ‘AaE’).

Keep the response sheets handy while you read the rest of this chapter, as we will often refer back to them. For now try to answer the following questions. Keep a record of your answers and come back to them when you have finished the chapter.

(i) Did your memories always contain both autobiographical *knowledge* as well as very detailed records of actual experiences?

(ii) Were the details always or predominantly in the form of visual mental images? If not, what form were they in?

(iii) Did you feel any emotions? Particularly with respect to the last two memories in comparison to the first memory.

(iv) Did the memories just ‘pop’ into your mind when you read the cue words, or did you have to elaborate the cue, for example, think about a chair at home and some incident associated with it, such as when you bought it?

(v) Did it take longer to retrieve a memory to cue two than to cues one and three?

(vi) How complete a record of the actual event would you say the memory is?

(vii) Did you notice how ‘time compressed’ the memory was? That is, it almost certainly took far less time to recall the event than the experience itself took.

(viii) How accurate, as a record of the experiences, were the memories?

(ix) At some point in the attempt to retrieve/construct the memories you must have decided that you had an appropriate memory in mind. Was this associated with any feelings? Did you have a sort of ‘Aha’ experience when the memory came to mind? Did you feel as though you were almost reliving at least a small part of the past (memory researchers call this *recollective experience*)?

equally important aspect is the *constructive* nature of autobiographical remembering. We know from the experience of our own memories that when knowledge of the past comes to mind, intentionally or spontaneously, it often features facts about ourselves and our lives, and images of people, locations, and activities, and, of course, detailed (episodic) memories of specific events may be recalled (as in the cue word experiment you have just completed). It is this coming together of conceptual autobiographical knowledge, generic images, and episodic memories that is the major form of construction in autobiographical remembering. In this section the nature and organization of autobiographical knowledge in long-term memory is considered first, followed by an account of episodic memories. The role of the self in memory construction is then reviewed, and finally the process of memory construction itself is outlined.

3.1 Autobiographical knowledge

Conway and Pleydell-Pearce’s (2000) model of autobiographical memory construction is essentially a hierarchy, with an interaction between knowledge and the records of specific one-off instances. Two broad types of autobiographical knowledge have been identified: general events and lifetime periods (Conway and Pleydell-Pearce, 2000). They vary according to how specific to actual events they are, and therefore where they sit in the hierarchy. Lifetime periods are less specific, and are made up of general events, that in turn are made up of specific events derived from actual experience (or episodic memories; see below and also see Chapter 8).

3.1.1 General events

General events refer to a variety of autobiographical knowledge structures such as single events (e.g. the day we went to South Wales), repeated events (e.g. Friday afternoons at the pub), and extended events (e.g. the conference in Sydney). General events may themselves be organized in several different ways. For example, they may be 'mini-histories' structured around detailed and sometimes vivid episodic memories of goal-attainment in developing skills, knowledge, and personal relationships (Robinson, 1992). Some general events may be of experiences of particular significance for the self and act as reference points for other associated general events (Pillemer, 1998; Singer and Salovey, 1993). Other general events may be grouped together because of their emotional similarity (McAdams *et al.*, 2001), and it is likely that there are yet other forms of organization at this level that await investigation (Brown and Schopflocher, 1998). However, the research currently available indicates that organization of autobiographical knowledge at the level of general events is extensive, and it appears to virtually always refer to progress in the attainment of highly self-relevant goals.

Conway and Pleydell-Pearce (2000) indicate that general events contain knowledge about locations, others, activities, feelings, and goals common to an event, as well as some specific episodic memories that help organize the general event knowledge. This autobiographical knowledge may be represented in several different ways and consist of a mental construction of images, feelings, and verbal statements (cf. Johnson-Laird, 1983 and Chapter 12). However, autobiographical knowledge in general events predominantly takes the form of generic visual images, i.e. images derived from repeated experiences (Brewer, 1986, 1988, 1996; Conway, 1996, 2001; Rubin and Greenberg, 1998). General event autobiographical knowledge can then be used to access associated sensory–perceptual episodic memories and, when it is used in this way, a specific and detailed autobiographical memory (AM) can be formed. Thus, a specific AM will usually, if not always, contain some general event knowledge and this will often be in the form of generic images (was this the case with the memories you recalled earlier?).

In one of the few studies of this type of knowledge, Robinson (1992) examined people's memories of the acquisition of skills (e.g. riding a bicycle or driving a car) and of aspects of personal relationships. These general events were found to be organized around a series of vivid memories relating to **goal attainment**. Consider two examples from Robinson's study:

> 1 *'Ever agreeable, and eager to do anything that would get me out of the doldrums of inferiority, my father rented a bike and undertook to help me to learn. . . . I shall always remember those first few glorious seconds when I realized I was riding on my own . . .'*
>
> (Quinn, 1990, cited in Robinson, 1992, p.224)

> 2 *'The first time I flew an airplane was one of the best firsts. It marked a sense of accomplishment for myself, and it also started me on the career path I have always wanted to follow. The day was warm and hazy, much as summer days in Louisville are. My nervousness didn't help the situation, as I perspired profusely. But as we took off from runway 6 the feeling of total euphoria took over, and I was no longer nervous or afraid. We cruised at 2,500 feet and I worked on some basic manoeuvres for approximately 45 minutes. We then returned to the airport, where I realized that this will soon be a career.'*
>
> (Robinson, 1992, p.226)

These 'first time' memories can cue other related memories and the whole general event carries powerful self-defining evaluations that persist over long periods of time. Importantly, Robinson found that many memories featured goal-related evaluative knowledge or self-defining memories (Singer and Salovey, 1993) along with more general knowledge and specific episodic memories. General event autobiographical knowledge can also be used to access related **lifetime periods** that contain associated knowledge.

3.1.2 Lifetime periods

Lifetime periods, like general events, contain representations of locations, others, activities, feelings, and goals common to the period they represent. They effectively encapsulate a period in memory and in so doing may provide ways in which access to autobiographical knowledge can be limited, channelled, or directed. As with general events there is evidence that lifetime periods contain evaluative knowledge, negative and positive, of progress in goal attainment (Beike and Landoll, 2000), and it seems likely that lifetime periods may play an important role in what Bluck and Habermas (2000) call the life story.

A life story is some more or less coherent theme or set of themes that characterize, identify, and give meaning to a whole life. A life story consists of several life story schema, which associate together selective autobiographical knowledge to define a theme (Bluck and Habermas, 2000). A schema is a memory structure that encapsulates an event such that common parts are fixed, while variable parts occur as 'slots'. Thus a schema for 'going to the cinema' would have pre-defined common parts (such as queuing for tickets, buying popcorn) and slots for variable parts (which cinema we went to, who I was with, what film we saw). Lifetime periods might provide the autobiographical knowledge that can be used to form life story schema and thus support the generation of themes – an idea that is very similar to Rubin and Berntsen's (2003) life scripts theory (presented above).

Lifetime period evaluations access related general events and, in turn, episodic memories that, when formed, provide the 'evidence' justifying the evaluations (see Beike and Landoll, 2000, and Conway and Pleydell-Pearce, 2000, for more on how autobiographical knowledge 'grounds' the self in memories of experience). In summary, lifetime period autobiographical knowledge is less event-specific than general event autobiographical knowledge; it is also more conceptual and abstract. It encapsulates significant parts of the life story and may form an important bridge from autobiographical memory to core aspects of the self, most likely through the generation of narrative 'scripts' of one's past experiences. Figure 17.2 depicts this scheme of autobiographical knowledge organization, and shows how such knowledge may be represented at different levels to form hierarchical partonomic knowledge structures. Partonomic refers to the way that a specific episodic memory is *part of* a general event, which in turn is *part of* a lifetime period, which is *part of* a life schema (Conway, 1996).

Recent research has aimed to provide some empirical support for this partonomic structure, using research into healthy older adults, who are known to have subtle deficits in memory. In this work, Piolino and colleagues (2010) have developed an autobiographical memory test to examine four levels of Conway and Pleydell-Pearce's model: lifetime periods, general events, specific events, and details. Participants have to generate as many different lifetime periods as possible in a given time period. For one of these lifetime periods, they then generate as many general events as they can retrieve, and for one general event, as many specific events, and so on. The evidence for the model comes from the fact that older adults, who have difficulties with episodic retrieval, struggle to generate the more specific levels of the hierarchy. This points to a critical difference between episodic and semantic memory. But you could try this yourself – you will probably find that the list of lifetime periods is somewhat easier to generate than specific events – the task gets more difficult the more you drill down through the hierarchy.

3.2 Episodic and semantic memory

In Tulving's (1972) original distinction between episodic and semantic memory, the defining feature of episodic memory was that it contained spatio-temporal information (see Chapter 8, Section 3.2). Episodic memories were of specific events that occurred at unique times, while semantic knowledge was of abstract, conceptual, context-free knowledge not linked to any specific place, time, or learning episode. The reference for episodic memory was then individual personal experience, whereas the reference for semantic knowledge was social and universal (Tulving, 1983). So, for example, if you think about what you ate for breakfast this morning, and retrieve vivid and specific details of this one experience, that is episodic memory. In

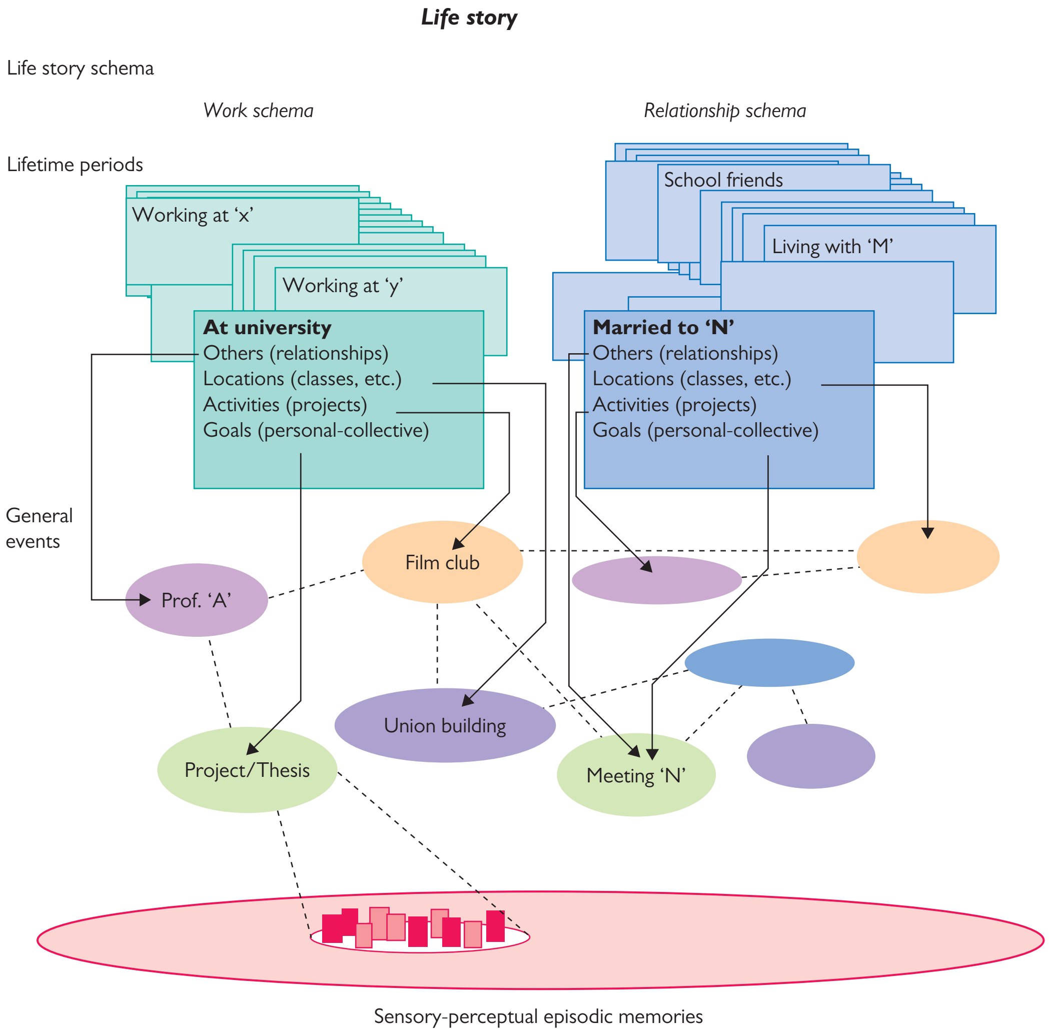

Life story
Life story schema
Work schema
Relationship schema
Lifetime periods
Working at ‘x’
Working at ‘y’
School friends
Living with ‘M’
At university
Others (relationships)
Locations (classes, etc.)
Activities (projects)
Goals (personal-collective)
Married to ‘N’
Others (relationships)
Locations (classes, etc.)
Activities (projects)
Goals (personal-collective)
General events
Prof. ‘A’
Film club
Union building
Project/Thesis
Meeting ‘N’
Sensory-perceptual episodic memories

FIGURE 17.2 Autobiographical knowledge and episodic memories.

comparison, recalling that the French traditionally have coffee and croissants for breakfast is general knowledge, reproduced from semantic memory.

In autobiographical memory, we see the interaction of these two forms of memory – and it would appear that they are not as clear-cut as these examples might suggest. For instance, episodic memories must, of course, contain semantic knowledge and this raises the question of how easily the two types of knowledge can be distinguished. Secondly, there are knowledge representations in long-term memory that on Tulving’s (1972) original definition we would want to term ‘semantic’, but which contain spatio-temporal information. For example, there arguably exists a schema for ‘breakfast’, which specifies the location, time, actions, order of actions, and objects involved in a typical breakfast (Schank and Abelson, 1977). Is this a semantic or episodic representation? A third issue relates to autobiographical knowledge. For instance, a person may know that last year they took a holiday in Italy – no other information needs to be brought to mind. But the reference for this knowledge (namely the holiday) is both personal and universal and, moreover, it clearly contains spatio-temporal knowledge (see also Dritschel *et al.*, 1992). Perhaps a better way to classify episodic and semantic has been provided by Tulving in more recent work. He suggests that episodic memories can be described as ‘mental time travel' where the self is able to retrieve specifics of a past experience. When recalled, these specific memories give rise to *recollective experience*, the feeling of experiencing the past, and this does not occur when other types of long-term knowledge are brought to mind (e.g. Tulving, 2002).

We now turn to recollective experience in autobiographical memory. Conway (2001) suggested that knowledge contained in episodic memories is largely *sensory–perceptual* in nature. Figure 17.2 conveys this by depicting episodic memories in an undifferentiated pool of representations of sensory–perceptual experiences. Thus, episodic memory is a repository of 'experience-near', highly event-specific sensory–perceptual details of recent experiences; experiences that lasted for comparatively short periods of time (seconds, minutes, or at most hours). These sensory–perceptual episodic memories do not endure in memory unless they become linked to more permanent autobiographical memory knowledge structures. Conway (2001) argues that access to sensory–perceptual episodic memories is rapidly lost. This is because most episodic memories do not become linked to more stable and permanent autobiographical knowledge in long-term memory and, as a consequence, they rapidly decay and become permanently inaccessible. As a simple demonstration, cast your mind back over the events of today. They will be extremely detailed and numerous. If you try the same exercise, remembering today's events, in a day or so, or perhaps next week, few episodic memories will have been retained relative to the number available on the day of experience.

The theory is that when experience-near sensory–perceptual fragments are accessed during memory construction, this is associated with recollective experience and the sense of 'remembering', as opposed to just knowing. As a concept, recollection is supported by a large number of studies looking at recognition memory for word lists. A number of prominent researchers including Tulving and Conway converge on the idea that it is the sense or experience of the self in the past and is associated with images, feelings, and other details that come to mind during remembering (see Gardiner and Richardson-Klavehn, 2000, for a review).

3.3 The working self

Constructing an autobiographical memory is a complex form of cognition and has several effects on processing generally. One of the main theories is that the entire cognitive system enters what Tulving (1983) called **retrieval mode**. In retrieval mode, attention is directed inwards towards internal representations of knowledge, and conscious awareness becomes dominated by these representations. As a memory is formed, the rememberer's awareness becomes emotionally influenced by recollective experience and a powerful sense of the self in the past arises. The division of attention that then occurs gives rise to an attenuation of all other cognitive processes and, because of this, recall of AMs could, potentially, be highly dysfunctional in that current processing sequences would be disrupted. In extreme cases, such as in the involuntary and intrusive recall of prior trauma that is symptomatic of PTSD, autobiographical recall may be pathologically disruptive to everyday functioning (as in Memory 5 at the beginning of this chapter). The point is that constructing a specific and detailed AM is a major cognitive occurrence with consequences for all other types of processing. Memory construction has therefore to be controlled, and according to Conway and Pleydell-Pearce (2000), this is one of the main functions of their concept of the working self (see also Markus and Ruvolo, 1989).

3.3.1 Goals and the working self

The working self is conceived as a complex hierarchy of interconnected goals, all of which are in varying states of activation, but only some of which can enter consciousness (see an extended discussion of goals in relation to the ACT-R cognitive architecture in Chapter 19). The working self may also contain representations of at least some goal-related knowledge, e.g. lifetime periods, life schema, and life story or stories, as well as currently active models of the self. It is through the working self goal structure that episodic memories are formed and autobiographical knowledge is abstracted from experience. Thus, goal-related experience is prioritized in terms of encoding, consolidation, accessibility, and construction into specific, if transitory, autobiographical memories. Strong evidence exists showing that overall goal orientation of particular personality types acts to raise the accessibility of goal-related autobiographical knowledge and so facilitate their recall. This work has its origin in a seminal paper by Markus (1977) who found that people with a strong personality trait relating to the dependent–independent dimension showed preferential access to memories of experiences in which they had behaved in dependent or independent ways. In contrast, individuals within whom the dependent–independent dimension was weak did not have this memory bias.

These types of self-memory congruency effects have since been observed in several studies and most especially in the work of McAdams on power, intimacy, and generativity (McAdams, 1982, 1985, 2001; McAdams, *et al*., 1997). McAdams (1982) used the thematic apperception test (TAT; Murray, 1938, 1943) to assess non-conscious aspects of personality (McClelland *et al*., 1989) and categorized individuals into those with a strong intimacy motivation or, in contrast, with a distinctive power motivation. Content analysis of subsequently freely recalled memories of 'peak' and other experiences found that the intimacy motivation group recalled peak experiences with a preponderance of intimacy themes compared to individuals who scored lower on this motivation, who in turn showed no memory bias. Similarly, the power motivation group recalled peak experiences with strong themes of power and satisfaction. Interestingly, neither group showed biases in memories for more mundane, less emotional, less self-defining memories. These striking biases in memory availability by dominant motive type suggest that the goal structure of the working self makes highly available those aspects of the knowledge base that relate most directly to currently active goals.

Work by Woike and her colleagues has further established the connection between personality and memory (Woike, 1995; Woike *et al*., 1999). In the tradition of personality research deriving from Murray (1938) and McClelland (e.g. McClelland *et al*., 1989), Woike identified implicit and explicit motives in a group of people who then recorded memorable events over a period of 60 days. According to McClelland *et al.* (1989), implicit motives are evident in preferences for certain types of affective experience such as 'doing well' for achievement and 'feeling close' for intimacy, whereas explicit motives are present in social values and aspects of the self that can be introspected. A corollary of this view is that affective experiences should give rise to memories associated with implicit motives. Explicit motives, on the other hand, should lead to memories of less affective, routine experiences, more closely associated with self-description than with measures of implicit motives (i.e. TAT performance). This was exactly Woike's finding in both a diary study and in a laboratory-based autobiographical memory retrieval experiment. Woike *et al.* (1999) also investigated groups of individuals classified as 'agentic' (concerned with personal power, achievement, and independence) or as 'communion' (concerned with relationships, interdependence, and others). Agentic personality types are considered to structure knowledge in terms of 'differentiation' (the emphasis is on differences, separateness, and independence) whereas communal individuals, in contrast, structure knowledge in terms of 'integration' (the emphasis is on similarity, congruity, and interdependence). Across a series of studies, people with agentic self-focus were found to consistently recall emotional memories of events that involved issues of agency (mastery, humiliation) with their content structured in terms of differentiation. People with communal self-focus recalled emotion memories featuring others, often significant others, in acts of love and friendship, with the memory content structured in terms of integration. These findings clearly implicate the self in determining recall and lend further weight to the suggestion that the self influences access to sets of goal-related memories.

In an intriguing study, Pillemer *et al.* (1996; also Pillemer, 1998) investigated memory for specific educational episodes (e.g. Memory 4 at the beginning of this chapter). The initial impetus for this work was the observation that autobiographies often contain accounts of highly specific events that were 'turning points' (self-defining moments) for the individual and usually involved the adoption of a superordinate life goal that then determined much of the individual's later activities. Pillemer *et al.* (1996) found that students and alumni were frequently able to report, in detail, highly vivid memories of interactions with professors and other teachers that had profoundly influenced their academic interests and, sometimes, the whole of their lives. These were often events in which superordinate long-term goals were adopted by the individual, e.g. to become a chemist, a writer, etc. Consider the following account by a postgraduate mature student of her first undergraduate Shakespeare class:

> *'I was fascinated by the easy way the professor roamed through Shakespeare, by just the amount of knowledge he had. He seemed to know everything. In fact, after class, I asked him if he could identify a quote I had found about fencing, "Keep up your bright swords, for the dew will rust them." Immediately he said "Othello, Act 1 Scene 2, I believe." Which turned out to be exactly right. I wanted to know a body of literature that well. I'm still working on it.'*
>
> (Pillemer et al., 1996, p.330)

Further emphasizing the power of specific episodic events, recent work by Kuwabara and Pillemer (2010) shows that activating a positive memory of someone's time at university makes them more likely to donate money to their university. Such research is critical because it demonstrates the link between autobiographical memory and behaviour. Of course, not all self-defining moments are positive and Pillemer *et al.* (1996; and Pillemer, 1998) list several other memories of more negative educational experiences that led to a subject being dropped, negatively conceived as 'difficult' or 'boring', and, in some cases, the emergence of negative conceptions of self as a poor or incompetent learner.

Singer and Salovey (1993) provide one of the main statements on the relation between goals and memories. A major finding in their study was that memories associated with feelings of happiness and pride were strongly linked with goal attainment and the smooth running of personal plans (see also Sheldon and Elliot, 1999). In contrast, memories associated with feelings of sadness and anger were linked to the progressive failure to achieve goals. Singer and Salovey (1993) proposed that each individual has a set of self-defining memories that contain critical knowledge of progress on the attainment of long-term goals. Such goals, e.g. attaining independence, intimacy, mastery, and so on, may have been adopted as solutions to dominant self-discrepancies arising from childhood experiences (Strauman, 1996). Related to this, Thorne (1995) found that the content of memories freely recalled across the lifespan by 20-year-olds conformed to what she called 'developmental truths'. Thus, memories from childhood very frequently referred to situations in which the child wanted help, approval, and love, usually from the parents, whereas memories from late adolescence and early adulthood referred to events in which the rememberer wanted reciprocal love, was assertive, or helped another.

Conway's notion of a 'working self' consisting of an active complex goal hierarchy is therefore a useful way in which to understand the pattern of findings from the study of personality and autobiographical memory. The evidence points to a particular role for the working self, which is to modulate access to knowledge in long-term memory and to control what new information enters the knowledge base and which episodic memories are rehearsed and become integrated with long-term knowledge structures. The working self's control structures may make some long-term knowledge highly accessible, e.g. self-congruent knowledge, whereas other knowledge may be inhibited, e.g. self-incongruent knowledge.

3.4 Constructing autobiographical memories

Autobiographical memories can either be intentionally constructed or, alternatively, may come to mind without the formation of any specific intention to recall a memory. We refer to the former type of construction as *generative retrieval* and the latter type as *direct retrieval*.

Generative retrieval occurs when remembering is intentional and the knowledge base is iteratively sampled as a memory is effortfully constructed. During this protracted process an initial cue is used to probe the knowledge base and accessed knowledge is evaluated against a retrieval model generated by the self. If the constraints of the retrieval model are satisfied then a memory is formed, and the knowledge activated in the knowledge base (by the cue), together with associated goals of the self, form the autobiographical memory in that episode of remembering. Usually this process takes several or more cycles of access, evaluation, and cue elaboration, as a stable pattern of activated knowledge that meets the constraints of the retrieval model gradually emerges. For example, in attempting to construct an AM to a cue such as 'cinema', a rememberer might elaborate the cue into the question 'When did I go to the cinema a lot?' This cue might lead to access of the lifetime period 'when I was a student'. Lifetime period knowledge can then be used to access general events, which in turn access episodic memories, and in this way a specific and task-relevant AM is constructed. Perhaps you were aware of this process when recalling memories to the cues 'chair', 'illness', and 'summer' in Activity 17.2? Figure 17.3 lists two protocols collected from people recalling memories to cue words while saying aloud what was going through their minds. Perhaps you were aware of similar types of knowledge coming to mind when you recalled your memories? Figure 17.4 provides a diagrammatic illustration of generative and direct retrieval.

Although the process of generative retrieval may seem laborious and is certainly effortful (retrieval times to word cues usually average between five and eight seconds), it nevertheless may operate with high efficiency when the system is in retrieval mode and multiple memories are to be recalled. Such circumstances arise in a conversation with another person

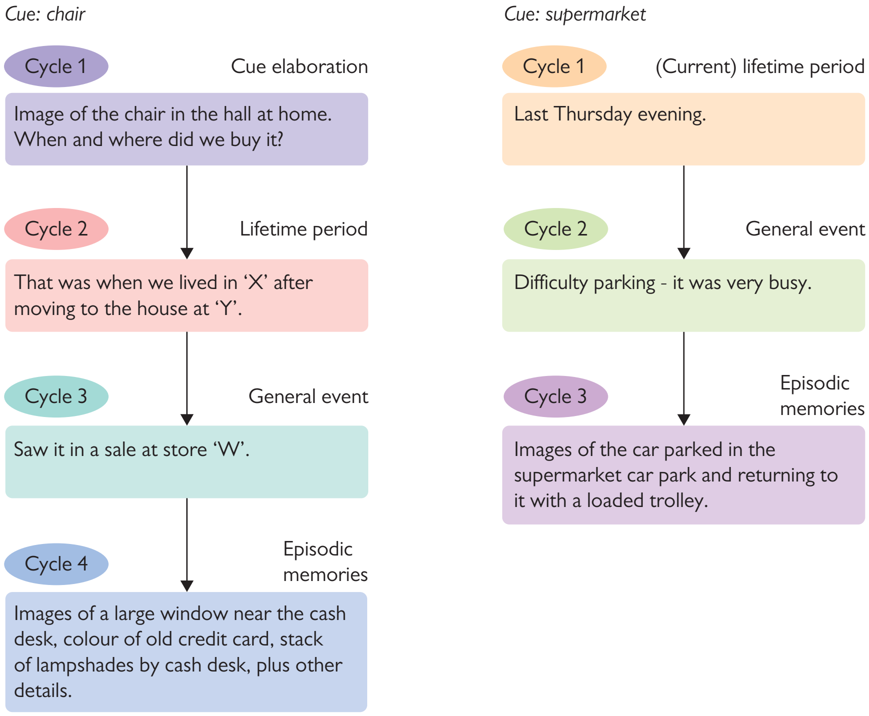

Cue: chair
Cycle 1
Cue elaboration
Image of the chair in the hall at home. When and where did we buy it?
Cycle 2
Lifetime period
That was when we lived in 'X' after moving to the house at 'Y'.
Cycle 3
General event
Saw it in a sale at store 'W'.
Cycle 4
Episodic memories
Images of a large window near the cash desk, colour of old credit card, stack of lampshades by cash desk, plus other details.
Cue: supermarket
Cycle 1
(Current) lifetime period
Last Thursday evening.
Cycle 2
General event
Difficulty parking - it was very busy.
Cycle 3
Episodic memories
Images of the car parked in the supermarket car park and returning to it with a loaded trolley.

FIGURE 17.3 Two protocols collected while rememberers related what came to mind when recalling memories to cue words.

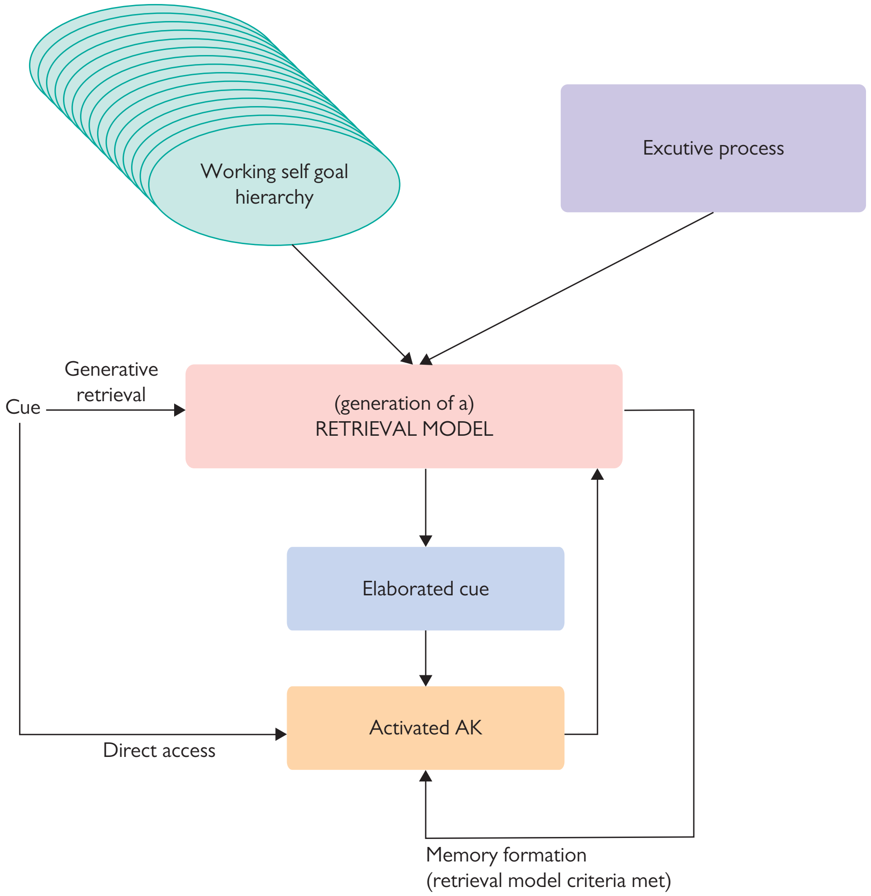


FIGURE 17.4 Direct and generative retrieval.

about a shared experience or in a discourse in which accounts of autobiographical memories are important. In conversation about the 'good old days' one does not feel held back by the effort of generating anecdotes or comparing notes on the specifics of a shared event.

Nevertheless, when a cue activates a general event and associated episodic memories, a specific autobiographical memory can be formed, apparently effortlessly and spontaneously. In certain instances, direct retrieval occurs. In direct retrieval a cue causes a pattern of activation in autobiographical knowledge (AK) that stabilizes as a specific autobiographical memory and bypasses the stages of generative retrieval (repeated autobiographical knowledge access, evaluation, and cue elaboration) (see Box 17.1). Sometimes, direct retrieval occurs completely involuntarily, when memories come to mind seemingly unbidden. Schlagman and Kvavilashvili (2008) conducted a systematic study of involuntary and voluntary autobiographical memories. They found that involuntary memories were more specific and faster retrieved than voluntary autobiographical memories. They were also more likely to be experienced in response to a negative cue. The study showed that involuntary memories often come to mind whilst engaged in repetitive and low-level cognitive tasks – in the case of their experiment, repeatedly detecting vertical lines amongst an array of horizontal lines. We return to direct access of negative events in the section on PTSD below. Compared to the research on voluntary autobiographical memory, there is a lack of data on involuntary memories. How often do you have the memory of a previous event come to mind whilst you are busy doing something else? Does it seem like a random occurrence, or might there also be a role for the self in what is retrieved involuntarily?

BOX 17.1 Haque and Conway's autobiographical memory 'probe' experiments

An experiment by Haque and Conway (2001) illustrates how both types of retrieval occur when people recall specific autobiographical memories in response to a range of cue words naming common activities, locations, and emotions. In this experiment the cue words were displayed on a computer screen and participants attempted to recall a memory for each cue individually.

In order to 'probe' the process of memory construction, a signal was displayed two seconds, five seconds, or 30 seconds after the cue word was on screen. In response to the signal the participant had to report as exactly as they could the current contents of consciousness.

The reports were then classified for the predominant type of knowledge they contained, i.e. lifetime period, general event, specific memory, or 'nothing in mind'. Table 17.2 shows the number and percentage of each type of report at each of the probe intervals. From Table 17.2 it can be seen that similar numbers of reports at the two-second probe contain either autobiographical knowledge (lifetime periods and general events) or specific autobiographical memories.

Autobiographical knowledge indicates the operation of the generative retrieval process (a memory has not yet been formed) whereas the report of specific autobiographical memories at this very short probe indicates direct retrieval. Thus both types of retrieval can occur in the same individual. As can also be seen from Table 17.2, the incidence of reports of autobiographical knowledge at the longer probe times sharply decreases (although note the persistence of some general event knowledge), while the formation of specific memories strongly increases. As might be expected, as the retrieval time lengthens so the generative process runs its course and specific autobiographical memories were formed.

TABLE 17.2 Distribution of protocols by protocol type and probe time in experiment 1 from Haque and Conway (2001)

Protocol type	**Time of probe**		
	2 seconds	5 seconds	30 seconds
Lifetime period	30 (70%)	10 (23%)	3 (6%)
General event	18 (32%)	23 (41%)	15 (27%)
Specific memory	33 (19%)	59 (35%)	79 (46%)
Nothing in mind	19 (63%)	8 (27%)	3 (10%)

SUMMARY OF SECTION 3

- Autobiographical remembering of a specific episode is *constructive* in nature. It brings together autobiographical knowledge (general events and lifetime periods), generic images, and episodic memories.
- In autobiographical memory, episodic memory is thought to consist of 'experience near', highly specific, sensory–perceptual details of recent experiences. Only those episodic memories that then go on to be linked to long-term memory will be available later to support the formation of subsequent autobiographical memories. It is helpful to think of episodic memories not by their content, but by the experience of recollection during retrieval.
- One role of the self is to control the process of autobiographical memory construction. This is because constructing an autobiographical memory is a major cognitive occurrence, and has consequences for all other types of processing. The working self can be thought of as a complex goal

hierarchy, modulating access to knowledge in long-term memory, and controlling knowledge that can enter the knowledge base. Episodic memories are then interpreted in terms of the self.

- The process of memory construction itself can be either intentional (generative retrieval) or unintentional (direct retrieval).

4 AUTOBIOGRAPHICAL MEMORY IN DISTRESS

As mentioned earlier (see Memory 5 at the beginning of this chapter) probably the most outstanding form of direct retrieval occurs in post-traumatic stress disorder, PTSD, in which memories for traumatic experiences figure prominently (Brewin and Holmes, 2003). In PTSD a range of symptoms are present but one of the most marked is that of persistent intrusive thoughts and memories. Consider the case study in Box 17.2.

Although autobiographical remembering usually involves an effortful and constructive process, in an individual like John it seems almost impossible to prevent memories of the trauma coming spontaneously to mind. Such direct retrieval of scenes from highly negative experiences is perhaps not so uncommon. Think back to a traumatic experience that you may have had, such as a car crash. Did it haunt you afterwards, with vivid images of the experience just 'popping' into mind? Or have you ever been to see a horror film and then found the next day that images of the worst scenes intrude into your mind? People with PTSD like John, and like the man who provided Memory 5 (at the beginning of the chapter) during therapy for PTSD relating to his road traffic accident, may have numerous intrusive memories of a trauma in a single day. Often these cause destabilizing emotions like intense anxiety, guilt, fear, and often all of these occur when a trauma image intrudes uncontrollably into consciousness. Memory intrusions in PTSD are highly disruptive to other cognitive processes; they hijack attention and ongoing experience, and in so doing make even the most routine tasks difficult. The sufferer is constantly being thrown into 'retrieval mode' and this diverts attention away from other processes. It is perhaps one of the reasons that John withdrew from work and daily life: the goals of everyday cognition were just too difficult for him to attain while his mental capacity was taken up with intrusive memories.

BOX 17.2 CASE STUDY John

(Details of the case have been changed in order to protect anonymity.)

'John' was seen for an emergency appointment with a psychiatrist due to his recent suicidal thoughts. The psychiatrist noted that three months previously he had seen a friend fall to his death from a building, but would not talk about it. He was then referred for assessment with a clinical psychologist. At the first meeting he appeared distracted, jumpy, and low in mood. He had stopped work three months previously and spent all day at home. He did not like to leave his house, although he felt safer going out in the dark. He could not listen to music. He reported feeling very tired as he frequently had nightmares. He also described being overwhelmed by mental images of his friend's death, which he tried hard to push away from his mind. He had periods where he felt unreal and cut off from other people.

PTSD is made up of several components (American Psychiatric Association, 1994): the traumatic event; response at the time of trauma; and subsequent psychological symptoms.

4.1 Traumatic event

First of all, the patient needs to report having experienced a 'traumatic event'. This is typically a situation in which the individual experienced or witnessed actual or threatened death, or serious injury to self or others. Examples include natural disasters, sexual assault, road traffic accidents, physical attack, and torture. In John's case, the trauma was seeing his friend fall down the centre shaft of a stairwell in a block of flats to his death.

4.2 Response at the time of trauma

It is not only the experience of a trauma that contributes to a diagnosis of PTSD. The person's reaction to the trauma is also critical. For example, soldiers fighting or doctors performing surgery are exposed to death or serious injury routinely. While many people in such occupations do at some stage become traumatized, many do not. In comparison to the number of people who have experienced 'a trauma', community studies have estimated variable rates of PTSD occurring within the general population ranging from 1 per cent to 14 per cent (American Psychiatric Association, 1994). Clearly, not everyone who has a trauma goes on to develop PTSD. Thus, the second component of PTSD is that the person's response to the trauma involved intense fear, hopelessness, or horror (American Psychiatric Association, 1994). For example, if there were two people in a car that crashed into the side of a bus, narrowly missing a head-on collision, one person might think 'I'm going to die' and be intensely afraid. The other person might feel only mild fear and conclude 'Phew! I'm so lucky this isn't worse'. Only the first person would display a symptom of PTSD and probably only this person would go on to develop the full range of PTSD symptoms. In other words the traumatic event has to be experienced as stressful in a major way and possibly a biological stress response must also occur if PTSD is to follow.

While the American Psychiatric Association's definition focuses on reactions of intense fear, hopelessness, and horror, other intense emotions are also frequently experienced at the time of trauma, such as anger or shame (Grey *et al.*, 2001). Conway and Pleydell-Pearce (2000) suggest that the perception of extreme danger and/or imminent death poses a fundamental challenge to the goal system of the working self. The trauma experience cannot be easily processed by the working self as it threatens the entire goal system, and because of this the experience cannot readily be integrated with autobiographical memory knowledge structures in long-term memory. On the other hand, if the working self were to survive such trauma it would be highly useful from a survival perspective to retain a detailed record of what occurred. Thus, the tension is between either not encoding the (life) threatening experience (traumatic amnesia) or keeping a good record of it just in case one survives (a vivid, or 'flashbulb', memory). In fact, what is frequently observed in PTSD is, initially at least, a fragmentary and often jumbled memory containing highly vivid details that are unordered with respect to their original order of experience, and often interspersed with extensive 'gaps' (islands of amnesia). This is perhaps just the sort of compromise we might expect when the working self is caught between encoding and not encoding.

In John's trauma, he was on the stairs several floors below his friend when he slipped, so could not see him initially. John's first reaction was intense fear that the banging sound meant impending danger. Then he experienced intense horror as he saw his friend fall past. He therefore meets the 'emotional response' criteria for PTSD. At the time John also felt intensely unreal, as if he was watching the event happen to him from outside of his body. He felt as if time had slowed down and that he was watching the event in slow motion. This phenomenon is known as detachment or dissociation (American Psychiatric Association, 1994), and is a common reaction during trauma. At the time of trauma, the feelings of unreality John experienced can be thought of as protecting the self from the destabilizing psychological impact of the trauma (van der Kolk *et al.*, 1996). That is, through mentally distancing himself from the situation, he may have protected the self from being overwhelmed with emotion. Interestingly, it is also worth noting that while dissociation may be protective at the time of trauma, it is also a strong predictor of developing PTSD (Shalev *et al.*, 1996).

4.3 Subsequent psychological symptoms

PTSD includes various psychological symptoms displayed by the patient. The symptoms present in three groups as outlined below.

4.3.1 Re-experiencing symptoms including intrusive memories

'Re-experiencing' a trauma includes having recurrent and intrusive recollections of the event (known collectively as 'intrusions'), recurrent distressing dreams, and 'flashbacks' that involve suddenly acting or feeling as if the event were happening again, as well as intense physiological reactivity and psychological distress to reminders of the event. Ehlers *et al.* (2004) provide many case study examples of intrusive flashback memories; one neatly illustrates just how powerful these traumatic episodic memories can be:

> *'A patient who thought that he was going to die during an assault and would never see his children again, was not able to access the fact that he actually survived and saw his children again when he remembered this particularly distressing point of the assault. And when the intrusion occurred he would again be overwhelmed with sadness.'*
>
> (Ehlers et al., 2004)

It should be noted that these sensory–perceptual–affective details are just the sort of autobiographical information contained in recently formed episodic memories in Conway's model outlined above. Indeed, it is central to the PTSD illness that the re-experience is mediated by episodic memories of the traumatic event itself. This stands in sharp contrast to other psychological disorders that may also feature intrusive imagery of an *imagined* traumatic event (as in some cases of psychosis, e.g. of being cut in two by a man wielding a large sword; Morrison *et al.*, 2002) or of a catastrophic future event (as in obsessive–compulsive disorder, e.g. violently attacking elderly parents with an axe; de Silva, 1986). Thus, John, for example, was plagued by countless episodic images of his trauma, and mentioning the incident in his assessment interview caused him to re-experience them. He had five specific images that intruded. These were:

- a banging noise
- seeing his friend fall past him
- the bottom of a helicopter ambulance
- the exterior of the block of flats
- himself swearing at another friend.

These episodic images contained sensory and emotional experience from the time of encoding that included mood states at the time of trauma (for John, fear and horror) and feelings of unreality.

4.3.2 Avoidance symptoms

Various reminders of the trauma trigger the re-experiencing of traumatic memories. In the exercise you did earlier you saw how cue words facilitate retrieval of detailed episodic memories. However, the key point about both generative and direct retrieval is that at some stage during retrieval a cue has to be present that can access the *content* of the sought-for knowledge (Tulving and Thompson, 1973). In PTSD, however, there may be many cues for the trauma knowledge (especially if this is represented in memory in terms of goals active at the time of the experience) and these have a tendency to generalize to other stimuli with shared features. For example, for a person raped by a bearded man, all men with beards may trigger re-experiencing symptoms, i.e. memory intrusions. Even more generally, for someone who had a road traffic accident with a red car, the colour red, even on post boxes or clothes, may trigger intrusive memories of the crash. These generalizations are made non-consciously and at first it may not be evident why an intrusive memory comes to mind. Ehlers *et al.* (2004) give the following example: 'A rape victim noticed that she was feeling extremely anxious while talking to a female friend in a restaurant and subsequently realized that the feeling was probably triggered by the presence of a man on another table who bore some physical similarity with the rapist' (Ehlers *et al.*, 2004).

PTSD sufferers rapidly learn what triggers their re-experiencing of intrusive memories and, once learned, such potent cues are avoided, which can sometimes lead to dysfunctional behaviour, e.g. avoiding all red objects of a certain size or leaving a restaurant abruptly. **Avoidance** is then the second cluster of PTSD symptoms. John, for example, stopped playing all music as any rhythmic beat caused him to re-experience the 'banging' image and distress. He stopped going out during the day as the sight of tall buildings also brought back powerful intrusions. He also avoided talking about the trauma. Although avoidance may feel helpful in the short term, in the longer term avoidance of reminders of a trauma will not enable someone with PTSD to recover. Avoidance forms part of the vicious cycle that maintains the disorder (Ehlers and Clark, 2000).

4.3.3 Amnesia as avoidance

A further form of avoidance is involuntary in nature and takes the form of amnesia for the trauma. This may occur because the self inhibits knowledge that was nonetheless encoded: often in therapy some memory returns to PTSD patients, suggesting that it was in fact inhibited rather than not encoded. On the other hand, the overwhelming of the working self by negative emotions may render encoding through this structure ineffective. The result is an amnesia more like that seen in anterograde amnesia following brain damage. This could also occur because the stress response causes a temporary increase in a particular neurohormone in a part of the brain, the medial temporal lobe, which as a result is temporarily disabled. At his therapy assessment John was not able to recall his memory of the time between seeing his friend fall past him at the top of the stairwell, and then when he was standing outside the building in a crowd. He said it felt like a 'gap' and, unlike the rest of his trauma, he was unable to recall it even with effort.

4.3.4 Hyperarousal symptoms

Hyperarousal may feel like being in a constant state of 'red alert' for potential danger. People with PTSD have an exaggerated startle response in that even small, unexpected noises make them jump. John repeatedly flinched throughout his assessment interview, for example if he heard a sound in the corridor outside. Other symptoms of hyperarousal are impaired concentration and irritability. Some patients are no longer able to concentrate on simple activities such as reading a newspaper or cooking. Impaired concentration links with features of autobiographical memory discussed earlier, that is, the process of retrieval of specific events (such as intrusions) is disruptive to other cognitive processes. John found it difficult to read the psychological assessment questionnaires and even to sustain attention when watching television.

4.4 Impact of symptoms

The final component of PTSD is the duration and impact of the disorder. To meet the diagnostic criteria the person must have had the cluster of PTSD symptoms for at least one month. This is because after a trauma most people typically have trauma symptoms, such as intrusive memories, for a short time. The impact criterion is standard to most mental disorders; that is, that the symptoms have caused significant and persistent distress for the person, for example in their occupation or socially. John had stopped working and also avoided all his friends. Indeed, he was so distressed by his symptoms and the impact that they had had on his life that he had begun to contemplate committing suicide.

In summary, then, to receive a diagnosis of PTSD, a patient must have experienced a traumatic event and responded to it with fear, helplessness, or horror. The patient must persistently re-experience the trauma, avoid stimuli associated with it, and have symptoms of increased arousal such as an exaggerated startle response. The trauma in John's case of PTSD was witnessing the violent death of a friend. John showed the full range of PTSD symptoms, especially intense intrusive memories, and these had a profound negative impact on his quality of life. Now we have gone through the diagnosis, go back and read the case vignette at the start of this section. Can you see how the symptoms John presented at assessment make sense?

4.5 The nature of intrusive trauma memories

Re-experiencing trauma in the form of intrusive memories is the hallmark symptom of PTSD, and because of it, PTSD is one of the major psychopathological disorders of autobiographical memory. Phenomenologically, intrusive trauma memories have several distinctive features that relate to our understanding of normal, non-traumatized autobiographical remembering. For instance, intrusive memories are image-based and they very often take the form of visual sensory–perceptual snapshots like 'film clips' (Ehlers and Steil, 1995) or episodic memories. The intrusive trauma images, although typically visual, may also incorporate sounds and smells and, sometimes, bodily sensations (Ehlers *et al.*, 2002). For example, a woman who was raped in the dark encoded memory in non-visual sensory modalities, and during treatment reported suddenly experiencing physical pain and smells. John's images were a mixture of visual and auditory images. The memories also contain the emotion experienced at the time of trauma, such as fear, shame, or disgust (Grey *et al.*, 2002), or feelings of unreality. As his friend fell past him, John felt unreal and saw the scene as if he were outside of his own body looking down on himself. He therefore experienced most of his intrusive images as if he was 'out-of-body' and often felt unreal. Also, intrusive memories can include verbal cognition from the time of trauma, typically catastrophic thoughts such as 'I'm going to die' (Holmes *et al.*, 2004).

Trauma memories have a quality of 'nowness' or 'live feel' (cf. Brown and Kulik, 1977). In normal autobiographical remembering the recall of specific events is accompanied by visual images and recollective experience, but in PTSD intrusive memories can be so compelling that the trauma feels as if it is in reality happening again. That is to say that the recollective experience component, the sense of the self in the past, appears to be overwhelmed or blocked by the intensity of re-experience: the PTSD sufferer does not have a sense or feeling of the self in the past; instead they are *actually* in that past moment. One woman with PTSD, for example, had a traumatic experience involving gun shots. Whenever she heard a sudden bang, such as a balloon pop at a children's birthday party, she would feel as if it was happening again and throw herself to the floor in protection. This type of experience is an example of a 'flashback'. Such cue-driven direct recall of overwhelming re-experienced memories is unusual in normal autobiographical remembering. Nevertheless, the same mechanisms may be operating: namely a cue accesses the content of an episodic memory and this becomes rapidly available to attention and consciousness. Without the self to effectively intervene in direct retrieval, the memory will capture attention and dominate consciousness. Perhaps this would occur much more frequently in normal recall if control processes did not act to prevent thoughts and memories coming to mind. Additionally, if episodic memories were more integrated with the autobiographical knowledge base it seems likely that they would not come to mind with such a feeling of 'nowness' and instead be recollectively experienced as a part of an extended and integral past. Indeed, one of the goals of successful treatment of PTSD is to reach a point at which, when a patient recalls a trauma memory, it is experienced more like a normal autobiographical memory, as a part of the past and less as a part of the 'now' (Ehlers and Clark, 2000).

Another point of departure between normal and trauma memories is that, unlike usual autobiographical remembering, the intrusive memories of trauma in people with PTSD often seem to be exact copies of what was experienced at the time of trauma. Moreover, the intrusions are usually highly consistent, being the same each time they come to mind. Such impressive consistency suggests that the same mental representation (episodic) memory is accessed each time an intrusion occurs. However, the veridicality of trauma images is a contentious issue and these may not always be based on experience. Holmes *et al.* (2004) found that, of a sample of patients with PTSD, approximately 2 per cent of different intrusive images were reported by participants as not actually being of their trauma experience. While this indicates that participants believed that most of their intrusions were of the event, it is possible that objectively they may not have been. Images, for instance, can be associated with beliefs that do not accurately reflect what in reality happened. Hackman *et al.* (2004) report the case of a woman who, after a house fire, experienced repeated intrusions of curtains burning. These led her to believe that her daughter was burning alive. However, she also had another intrusion of when she saw her daughter's body in the morgue, which was unburned. The daughter had in fact died of smoke inhalation. The patient was, however, unable to connect the different information in the two images. Possibly these contradictory, highly vivid trauma images reflect unresolved affective conflicts in the patient.

Incidents of distortion and false images in trauma memories do then occur (Ehlers *et al.*, 2004; Conway *et al.*, 2004) and they pose an interesting question: if a memory of a trauma is created in part to preserve a detailed record of what occurred when one survived a trauma, how can it contain distortions and errors? The question has not yet been addressed by the appropriate research but we might speculatively consider the role of the self in generating fantasies that protect the self from deeply undermining cognition. Whatever eventually turns out to be the explanation, it is clear that distortion in PTSD images requires just as much explanation as do accurate flashbacks.

Trauma images are like 'highlights' picked out in a trailer for a film, except that they can occur one by one rather than as a sequence and often do not appear to coalesce in any obvious way, instead presenting as fragmented and disorganized (Foa *et al.*, 1995). Earlier it was reported that John had five distinct images of his trauma and these too appeared unorganized and fragmented. Interestingly, most PTSD patients spontaneously report between three to five trauma images (Hackmann *et al.*, 2004) suggesting that there may be some consistency in accessing trauma memories. More generally, however, one of the fascinating questions about trauma memories is why are there intrusive images of some moments but not others? Researchers have begun to address this question by investigating 'hotspots' in trauma memories (Richards and Lovell, 1999; Ehlers and Clark, 2000). These are elicited by asking the patient 'What are the worst parts of your trauma when you describe it?' These worst moments correspond to the moments that intrude. Grey *et al.* (2001) found that hotspot images are associated with a wide range of 'peak' emotions. They consist of the sensory–perceptual information encoded at that point in time, as well as the cognition linked to the specific emotion. For example, a patient had an extremely fear-filled image of the sound of impact during a crash and the sensation of being flung forwards, accompanied by the cognition 'I'm going to die'. Hotspots that return as intrusive memories may then relate to moments during the experience of trauma when the self was most intensely challenged and, clearly, repetitively and intrusively recalling such moments must act to destabilize the self.

SUMMARY OF SECTION 4

- Direct retrieval in an extreme, disruptive, and distressing form is illustrated by the intrusive memories of trauma (e.g. flashbacks) experienced by people with post-traumatic stress disorder (PTSD).
- The features of PTSD include the traumatic episode itself, the person's experience at encoding, symptoms of re-experiencing the trauma in memory, avoidance and amnesia, and hyperarousal. The case study considered the clinical features from an autobiographical memory perspective.
- Section 4 describes an application of theory to a clinical condition, where arguably memory research has helped inform the care and understanding of a relatively common condition. The keen student will find other instances of the application of theory – to brain damage, depression, child development, and autism, to name but a few. We have given some suggestions for further reading below.

5 WHAT NEXT FOR AUTOBIOGRAPHICAL MEMORY AND THE SELF?

According to the search engine 'the Web of Knowledge', 87 scientific articles were published with 'autobiographical memory' in the title in 2010. Ten years before, just 29 were published that year, and ten years before that, only nine were published. It is, then, an area of growth in psychology and a topic that is perhaps hard to do justice to in one chapter. We have skipped over much interesting work on autobiographical memory in the brain, for instance, so we will try to give a brief flavour for it here.

What has driven the growth in research is a need to understand one of the most complex aspects of human behaviour, and to understand how real life may be affected by memory impairment caused by brain damage, for instance. Perhaps surprisingly for a topic that started off as a niche subject within a very large field – human memory – autobiographical memory is now at the forefront of neuroscientific accounts of memory (e.g. Ryan *et al.*, 2001). Its popularity is undoubtedly due to the fact that autobiographical memory intersects with topics that are central to understanding the human condition, such as emotion, personality, and social interaction, and this is apparent in the work reviewed in this chapter.

As this chapter has shown, autobiographical memory represents an intricate coming together of separate stores and processes, and we have presented the view that what knits together and constrains this process is the self – or more specifically in the case of Conway's view, the working self. The working self, as envisaged as part of the self-memory system, is just one of several theoretical accounts that attempt to provide a framework for understanding and examining autobiographical memory.

Essentially, all of these alternative theories grapple with the same idea: that autobiographical memory is a construction from different stores and neurological regions, and that a controlling process is required to integrate these separate stores. For instance, Moscovitch (e.g. Moscovitch and Melo, 1997) provides a model of strategic retrieval from memory that requires coordination from the frontal lobes of the brain. This model operates at more of a neurological level than those presented above. As his data, Moscovitch uses neuropsychological studies which show that people with frontal lobe damage misremember and fabricate memories as a result of a quality control failure: memories become reconstructed in a manner that does not correspond to reality (see also Schacter *et al.*, 1998 and Baddeley and Wilson, 1988). These theories all rely on some higher order process to oversee the reconstruction and reanimation of memories into a meaningful narrative account.

Our own recent research has looked to provide some support for the notion that the self is involved in the organization of autobiographical memory. Our tactic was to collect participants' 'selves' and see how memories were distributed temporally around these. A sociologist, Kuhn (e.g. Kuhn and McPartland, 1954), developed a simple tool for researching the self. In his Twenty Statements Test, people have to complete the phrase 'I am . . .' with statements that reflect their self, such as *I am a psychologist, I am ginger, I am a husband, I am impatient*. We examined the characteristics of the memories generated for these selves when their 'I am' statements were given back as memory cues (Rathbone *et al.*, 2008), a paradigm we named the 'IAM Task' (I am memory task). This generated a pronounced clustering around the time of self formation. For instance, *I am a husband* might be supported with a set of memories from the time the participant got married. This was even the case for less concrete selves, such as *I am impatient*. Participants were able to generate specific memories for such cues, and the dates of the memories clustered around the time when they judged that impatience first became part of their identity.

This procedure then enabled the examination of the relationship between the self and memory in amnesia. In the case of PJM (see Box 17.3; Rathbone *et al.*, 2009), we showed that even someone with little or no episodic memory generated appropriate 'I am' statements, and with these a set of clustered events, even if these events were not remembered or experienced as episodic memories in the sense described above. Our interpretation of this was that the self is maintained even without episodic memory, and semantic knowledge about life events takes on the role of supporting the sense of self. We found similar results in a case of memory impairment in epilepsy (Illman *et al.*, 2011).

This chapter has discussed many studies that explore the link between self and memory, but perhaps you've found yourself wondering what this 'self' is, how it is defined, and how it is measured? These are all important questions, from both scientific and philosophical points of view. In order to study the self, we must first define it in some way. For example, in the studies that involve 'I am' statements, the self is examined in terms of aspects of self-knowledge (e.g. 'I am a researcher'). However, tapping into the more nebulous aspects of self, such as what it feels like to be 'me' from moment to moment, is a difficult endeavour. In Section 3 we discussed how the working self controls memory construction, but one might ask the question of what controls the working self? This philosophical problem is similar to the homunculus argument (Gregory, 1987), in that one apparently needs to keep searching one step back for a controlling agent that oversees conscious experience. Although no one has yet solved the puzzle of how to explain conscious experience (see Chapter 18), these difficult conceptual issues are part of what makes the topic of self and memory such an intriguing area of research.

It is perhaps fitting to finish up by looking to the future. The autobiographical memory system described in this chapter is clearly a complex, dynamic system, which is arguably shaped by personal goals and the need to orientate oneself in a life story. Perhaps unsurprisingly, one of the most recent developments in the field has been the notion that this reconstructive, goal-driven system is at play when looking into the future as well as the past (Schacter and Addis, 2007). According to this view, overlapping brain and cognitive systems are responsible for both imagining the future and retrieving the past. The evidence for their similarity comes largely from neuroimaging studies, with both autobiographical remembering and imagining the future seeming to rely on the

BOX 17.3 What happens to the self when memory fails?

PJM suffered a brain injury after falling from a bicycle, and this damage left her with an inability to remember episodic memories from several years of her life. Some key life events were missing from PJM's autobiographical memory, including moving house and having a baby – events that most people would have many salient and readily accessible memories for. Although PJM lacked episodic recollection, she was able to retrieve semantic facts about her past. Thus she *knew* she had moved house, but could not recall any specific events associated with that knowledge.

Rathbone *et al.* (2009) were interested in exploring how this change in PJM's autobiographical memory might affect her sense of self, and the way that memories were used to support self-knowledge in the IAM Task. Results showed that PJM's semantic memories clustered temporally around periods of identity development (e.g. when she became a mother), suggesting that semantic facts support information about the self in much the same fashion as episodic memories. Furthermore, PJM's ability to generate and date 'I am' statements was not impaired, indicating that the ability to possess information about the self is not reliant on episodic memory.

hippocampus, a small region of the brain traditionally thought to be responsible for memory and navigation (Addis *et al.*, 2010). Interestingly, the same issues, such as specificity and the relationship between episodic and semantic systems, emerge as research themes in future thinking just as in autobiographical memory. And it would appear that the self is critical in future thinking too – people's 'I will be . . .' statements are supported by imagined events that seem to cluster temporally in the same way that events and recollections are distributed around the self in the past (Rathbone *et al.*, 2011).

FURTHER READING

Conway, M.A. and Pleydell-Pearce, C.W. (2000) 'The construction of autobiographical memories in the self memory system', *Psychological Review*, vol.107, no.2, pp.261–88.

McAdams, D.P. (2001) 'The psychology of life stories', *Review of General Psychology*, vol.5, no.2, pp.100–22. These articles review research on autobiographical memory and propose, respectively, models in which memories are mental constructions within a self-memory system, and in which people construct narratives of the self.

Brewin, C.R. and Holmes, E.A. (2003) 'Psychological theories of posttraumatic stress disorder', *Clinical Psychology Review*, vol.23, no.3, pp.339–76. Brewin and Holmes review recent research on the psychological processes involved in PTSD and its treatment, describe early theoretical approaches, and compare three recent theories.

Addis, D.R. and Tippett, L. J. (2004) 'Memory of myself: autobiographical memory and identity in Alzheimer's disease', *Memory*, vol.12, pp.56–74. A recent study of the integrity of autobiographical memory and identity in Alzheimer's disease.

Kuwabara, K.J. and Pillemer, D.B. (2010) 'Memories of past episodes shape current intentions and decisions', *Memory*, vol.18, no.4, pp.365–74. A recent study that tested the idea that people's decisions and future plans are influenced by recently activated autobiographical memories.

Rathbone, C.J., Moulin, C.J.A., and Conway, M.A. (2009) 'Autobiographical memory and amnesia: using conceptual knowledge to ground the self', *Neurocase*, vol.15, no.5, pp.405–18. A study that reports on a patient with retrograde amnesia, PJM, and sheds light on the relationship between self, episodic memory, and autobiographical knowledge.

REFERENCES

Addis, D.R., Chen, T.P., Roberts, R., and Schacter, D.L. (2010) 'Hippocampal contributions to the episodic simulation of specific and general future events', *Hippocampus*, doi: 10.1002/hipo.20870.

American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV*, 4th edition, Washington, DC, American Psychiatric Association.

Baddeley, A.D. (1986) *Working Memory*, Oxford, Clarendon Press.

Baddeley, A.D. (2000) 'The episodic buffer: a new component of working memory?' *Trends in Cognitive Sciences*, vol.4, no.11, pp.417–23.

Baddeley, A. and Wilson, B. (1988) 'Frontal amnesia and the dysexecutive syndrome', *Brain & Cognition*, vol.7, no.2, pp.212–30.

Beike, D.R. and Landoll, S.L. (2000) 'Striving for a consistent life story: cognitive reactions to autobiographical memories', *Social Cognition*, vol.18, no.3, pp.292–318.

Berntsen, D. and Rubin, D.C. (2004) 'Cultural life scripts structure recall from autobiographical memory', *Memory & Cognition*, vol.32, pp.427–42.

Bluck, S. and Habermas, T. (2000) 'The life story schema', *Motivation and Emotion*, vol.24, no.2, pp.121–47.

Bohn, A. and Berntsen, D. (2011) 'The reminiscence bump reconsidered: children's prospective life stories show a bump in young adulthood', *Psychological Science*, vol.22, no.2, pp.197–202.

Brewer, W.F. (1986) 'What is autobiographical memory?' in Rubin, D.C. (ed.) *Autobiographical Memory*, Cambridge, Cambridge University Press.

Brewer, W.F. (1988) 'Memory for randomly sampled autobiographical events', in Neisser, U. and Winograd, E. (eds.) *Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory*, New York, Cambridge University Press.

Brewer, W.F. (1996) 'What is recollective memory?' in Rubin, D.C. (ed.) *Remembering Our Past: Studies in Autobiographical Memory*, Cambridge, Cambridge University Press.

Brewin, C.R. and Holmes, E.A. (2003) 'Psychological theories of posttraumatic stress disorder', *Clinical Psychology Review*, vol.23, no.3, pp.339–76.

Brown, N.R. and Schopflocher, D. (1998) 'Event cueing, event clusters, and the temporal distribution of autobiographical memories', *Applied Cognitive Psychology*, vol.12, no.4, pp.305–19.

Brown, R. and Kulik, J. (1977) 'Flashbulb memories', *Cognition*, vol.5, no.1, pp.73–99.

Conway, M.A. (1996) 'Autobiographical memories and autobiographical knowledge', in Rubin, D.C. (ed.) *Remembering Our Past: Studies in Autobiographical Memory*, Cambridge, Cambridge University Press.

Conway, M.A. (2001) 'Sensory perceptual episodic memory and its context: autobiographical memory', *Philosophical Transactions of the Royal Society, Series B, Biological Sciences*, vol.356, no.1413, pp.1375–84.

Conway, M.A. and Pleydell-Pearce, C.W. (2000) 'The construction of autobiographical memories in the self memory system', *Psychological Review*, vol.107, no.2, pp.261–88.

Conway, M.A., Meares, K., and Standart, S. (2004) 'Images and goals', *Memory*, vol.12, no.4, pp.525–31.

de Silva, P. (1986) 'Obsessional compulsive imagery', *Behaviour Research and Therapy*, vol.24, no.3, pp.333–50.

Dritschel, B., Williams, J.M.G, Baddeley, A.D., and Nimmo-Smith, I. (1992) 'Autobiographical fluency: a method for the study of personal memory', *Memory and Cognition*, vol.20, pp.133–40.

Ehlers, A. and Clark, D.M. (2000) 'A cognitive model of posttraumatic stress disorder', *Behaviour Research and Therapy*, vol.38, no.4, pp.319–45.

Ehlers, A. and Steil, R. (1995) 'Maintenance of intrusive memories in posttraumatic stress disorder: a cognitive approach', *Behavioural and Cognitive Psychotherapy*, vol.23, no.3, pp.217–49.

Ehlers, A., Hackmann, A., and Michael, T. (2004) 'Intrusive re-experiencing in post-traumatic stress disorder: phenomenology, theory and therapy', *Memory* vol.12, no.4, pp.303–415.

Ehlers, A., Hackmann, A., Steil, R., Clohessy, S., Wenninger, K., and Winter, H. (2002) 'The nature of intrusive memories after trauma: the warning signal hypothesis', *Behaviour Research and Therapy*, vol.40, no.9, pp.995–1002.

Erikson, E.H. and Erikson, J.M. (1982/1997) *The Life Cycle Completed*, New York, W.W. Norton & Co.

Fitzgerald, J.M. (1988) 'Vivid memories and the reminiscence phenomenon: the role of a self narrative', *Human Development*, vol.31, pp.261–73.

Fitzgerald, J.M. (1996) 'Intersecting meanings of reminiscence in adult development and aging', in Rubin, D.C. (ed.) *Remembering Our Past: Studies in Autobiographical Memory*, Cambridge, Cambridge University Press.

Fitzgerald, J.M. and Lawrence, R. (1984) 'Autobiographical memory across the life-span', *Journal of Gerontology*, vol.39, no.6, pp.692–8.

Fivush, R., Hammond, C., and Reese, E. (1996) 'Remembering, recounting, and reminiscing: the development of autobiographical memory in social context' in Rubin, D.C. (ed.) *Remembering Our Past: Studies in Autobiographical Memory*, Cambridge, Cambridge University Press.

Foa, E.B., Molnar, C., and Cashman, L. (1995) 'Change in rape narratives during exposure therapy for posttraumatic stress disorder', *Journal of Traumatic Stress*, vol.8, no.4, pp.675–90.

Franklin, H.C. and Holding, D.H. (1977) 'Personal memories at different ages', *Quarterly Journal of Experimental Psychology*, vol.29, pp.527–32.

Freud, S. (1955, first published 1899) 'Screen memories', in Strachey, J. (ed. and trans.) *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume 3*, London, Hogarth Press.

Fromholt, P. and Larsen, S.F. (1991) 'Autobiographical memory in normal aging and primary degenerative dementia (dementia of the Alzheimer type)', *Journal of Gerontology: Psychological Sciences*, vol.46, pp.85–91.

Fromholt, P. and Larsen, S.F. (1992) 'Autobiographical memory and life-history narratives in aging and dementia (Alzheimer type)', in Conway, M.A., Rubin, D.C., Spinnler, H., and Wagenaar, W. (eds.) *Theoretical Perspectives on Autobiographical Memory*, Utrecht, Kluwer Academic Publishers.

Gallagher, T., Campbell, M., and Gillies, M. (1995) *The Smiths: All Men Have Secrets*, London, Virgin Books.

Gardiner, J.M. and Richardson-Klavehn, A. (2000) 'Remembering and knowing', in Tulving, E. and Craik, F.I.M. (eds.) *Handbook of Memory*, Oxford, Oxford University Press.

Gilhooly, K.J. and Logie, R.H. (1980) 'Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words', *Behaviour Research Methods and Instrumentation*, vol.12, pp.395–427.

Gregory, R.L. (ed.) (1987) *The Oxford Companion to the Mind*, Oxford, Oxford University Press.

Grey, N., Holmes, E., and Brewin, C. (2001) 'It's not only fear: peri-traumatic emotional "hot spots" in posttraumatic stress disorder', *Behavioural and Cognitive Psychotherapy*, vol.29, no.3, pp.367–72.

Grey, N., Young, K., and Holmes, E. (2002) 'Cognitive restructuring within reliving: a treatment for peri-traumatic emotional "hot spots" in posttraumatic stress disorder', *Behavioural and Cognitive Psychotherapy*, vol.30, no.1, pp.37–56.

Hackmann, A., Ehlers, A., Speckens, A., and Clark, D.M. (2004) 'Characteristics and content of intrusive memories in PSTD and their changes with treatment', *Journal of Traumatic Stress*, vol.17, no.3, pp.231–40.

Haque, S. and Conway, M.A. (2001) 'Probing the process of autobiographical memory retrieval', *European Journal of Cognitive Psychology*, vol.13, no.3, pp.1–19.

Hollway, W. and Jefferson, T. (2000) *Doing Qualitative Research Differently: Free Association, Narrative and the Interview Method*, London, Sage.

Holmes, A. and Conway, M.A. (1999) 'Generation identity and the reminiscence bump: memories for public and private events', *Journal of Adult Development*, vol.6, pp.21–34.

Holmes, E.A., Grey, N., and Young, K.A.D. (2005) 'Intrusive images and "hotspots" of trauma memories in posttraumatic stress disorder: an exploratory investigation of emotions and cognitive themes', *Journal of Behaviour Therapy and Experimental Psychiatry*, vol.36, no.1, pp.3–17.

Howe, M.L. and Courage, M.L. (1997) 'The emergence and early development of autobiographical memory', *Psychological Review*, vol.104, pp.499–523.

Illman, N.A., Rathbone, C.J., Kemp, S., and Moulin, C.J.A. (2011) 'Autobiographical memory and the self in a case of transient epileptic amnesia', *Epilepsy & Behavior*, vol.21, no.1, 36–41.

Janssen, S.M.J., Rubin, D.C., and St. Jacques, P.L. (2011) 'The temporal distribution in autobiographical memory: changes in reliving and vividness over the life span do not explain the reminiscence bump', *Memory & Cognition*, vol.39, no.1, 1–11.

Johnson-Laird, P.N. (1983) *Mental Models*, Cambridge, MA, Harvard University Press.

Kuhn, M.H. and McPartland, T.S. (1954) 'An empirical investigation of self-attitudes', *American Sociological Review*, vol.19, pp.68–76.

Kuwabara, K.J. and Pillemer, D.B. (2010) 'Memories of past episodes shape current intentions and decisions', *Memory*, vol.18, no.4, 365–74.

Larsen, S.F. (1998) 'What is it like to remember? On phenomenal qualities of memory', in Thompson, C.P., Herrmann, D.J., Bruce, D., Reed, J.D., Payne, D.G., and Toglia, M.P. (eds.) *Autobiographical Memory: Theoretical and Applied Perspectives*, Mahwah, NJ, Erlbaum.

Markus, H. (1977) 'Self-schemata and processing information about the self', *Journal of Personality and Social Psychology*, vol.35, no.2, pp.63–78.

Markus, H. and Ruvolo, A. (1989) 'Possible selves: personalized representations of goals', in Pervin, L.A. (ed.) *Goal Concepts in Personality and Social Psychology*, Hillsdale, NJ, Lawrence Erlbaum Associates.

McAdams, D.P. (1982) 'Experiences of intimacy and power: relationships between social motives and autobiographical memory', *Journal of Personality and Social Psychology*, vol.42, no.2, pp.292–302.

McAdams, D.P. (1985) *Power, Intimacy, and the Life Story: Personological Inquiries into Identity*, New York, Guilford Press.

McAdams, D.P. (2001) 'The psychology of life stories', *Review of General Psychology*, vol.5, no.2, pp.100–22.

McAdams, D.P., Diamond, A., de Audin, E., and Mansfield (1997) 'Stories of commitment: the psychosocial construction of generative lives', *Journal of Personality and Social Psychology*, vol.72, no.3, pp.678–94.

McAdams, D.P., Reynolds, J., Lewis, M.L., Patten, A., and Bowman, P.T. (2001) 'When bad things turn good and good things turn bad: sequences of redemption and contamination in life narrative, and their relation to psychosocial adaptation in midlife adults and in students', *Personality and Social Psychology Bulletin*, vol.27, pp.472–83.

McClelland, D.C., Koestner, R., and Weinberger, J. (1989) 'How do self-attributed and implicit motives differ?' *Psychological Review*, vol.96, no.4, pp.690–702.

Morrison, A.P., Beck, A.T., Glentworth, D., Dunn, H., Reid, G., Larkin, W., and Williams, S. (2002) 'Imagery and psychotic symptoms: a preliminary investigation', *Behaviour Research and Therapy*, vol.40, pp.1063–72.

Morrison, C.M. and Conway, M.A. (2010) 'First words and first memories', *Cognition*, vol.116, no.1, pp.23–32.

Moscovitch, M. and Melo, B. (1997) 'Strategic retrieval and the frontal lobes: evidence from confabulation and amnesia', *Neuropsychologia*, vol.35, no.7, 1017–34.

Murray, H.A. (1938) *Explorations in Personality*, New York, Oxford University Press.

Murray, H.A. (1943) *The Thematic Apperception Test: Manual*, Cambridge, MA, Harvard University Press.

Pillemer, D.B. (1998) *Momentous Events, Vivid Memories*, Cambridge, MA, Harvard University Press.

Pillemer, D.B. and White, S.H. (1989) 'Childhood events recalled by children and adults', in Reese, H.W. (ed.) *Advances in Child Development and Behaviour, Volume 21*, San Diego, CA, Academic Press.

Pillemer, D.B., Picariello, M.L., Law, A.B., and Reichman, J.S. (1996) 'Memories of college: the importance of specific educational episodes', in Rubin, D.C. (ed.) *Remembering Our Past: Studies in Autobiographical Memory*, Cambridge, Cambridge University Press.

Piolino, P., Coste, C., Martinelli, P., Macé, A.L., Quinette, P., Guillery-Girard, B., and Belleville, S. (2010) 'Reduced specificity of autobiographical memory and aging: do the executive and feature binding functions of working memory have a role?' *Neuropsychologia*, vol.48, pp.429–40.

Rathbone, C.J., Conway, M.A., and Moulin, C.J.A. (2011) 'Remembering and imagining: the role of the self', *Consciousness & Cognition*, vol.20, pp.1175–82.

Rathbone, C.J., Moulin, C.J.A., and Conway, M.A. (2008) 'Self-centred memories: the reminiscence bump and the self', *Memory & Cognition*, vol.36, no.8, 1403–14.

Rathbone, C.J., Moulin, C.J.A., and Conway, M.A. (2009). 'Autobiographical memory and amnesia: using conceptual knowledge to ground the self', *Neurocase*, vol.15, no.5, pp.405–18.

Richards, D. and Lovell, K. (1999) 'Behavioural and cognitive interventions in the treatment of PTSD', in Yule, W. (ed.) *Post-traumatic Stress Disorders: Concepts and Therapy*, Chichester, Wiley.

Robinson, J.A. (1992) 'First experience memories: contexts and function in personal histories', in Conway, M.A., Rubin, D.C., Spinnler, H., and Wagenaar, W. (eds.) *Theoretical Perspectives on Autobiographical Memory*, Dordrecht, The Netherlands, Kluwer Academic Publishers.

Rubin, D.C. (2002) 'Autobiographical memory across the lifespan', in Graf, P. and Ohta, N. (eds.) *Lifespan Development of Human Memory*, Cambridge, MA, MIT Press.

Rubin, D.C. and Berntsen, D. (2003) 'Life scripts help to maintain autobiographical memories of highly positive, but not highly negative, events', *Memory & Cognition*, vol.31, pp.1–14.

Rubin, D.C. and Greenberg, D.L. (1998) 'Visual-memory-deficit amnesia: a distinct amnesic presentation and etiology', *Proceedings of the National Academy of Sciences*, vol.95, pp.1–4.

Rubin, D.C. and Schulkind, M.D. (1997) 'The distribution of autobiographical memories across the lifespan', *Memory and Cognition*, vol.25, pp.859–66.

Rubin, D.C., Rahhal, T.A., and Poon, L.W. (1998) 'Things learned in early adulthood are remembered best', *Memory and Cognition*, vol.26, pp.3–19.

Rubin, D.C., Wetzler, S.E., and Nebes, R.D. (1986) 'Autobiographical memory across the adult lifespan', in Rubin, D.C. (ed.) *Autobiographical Memory*, New York, Cambridge University Press.

Ryan, L., Nadel, L., Keil, K., Putnam, K., Schnyer, D., Trouard, T., and Moscovitch, M. (2001) 'Hippocampal complex and retrieval of recent and very remote autobiographical memories: evidence from functional magnetic resonance imaging in neurologically intact people', *Hippocampus*, vol.11, no.6, pp.707–14.

Schacter, D.L. and Addis, D.R. (2007) 'The cognitive neuroscience of constructive memory: remembering the past and imagining the future', *Philosophical Transactions of the Royal Society, Series B, Biological Sciences*, vol.362, no.1481, pp.773–86.

Schacter, D.L., Norman, K.A., and Koutstaal, W. (1998) 'The cognitive neuroscience of constructive memory', *Annual Review of Psychology*, vol.49, pp.289–318.

Schank, R.C. and Abelson, R.P. (1977) *Scripts, Plans, Goals, and Understanding*, Hillsdale, NJ, Erlbaum.

Schlagman, S. and Kvavilashvili, L. (2008) 'Involuntary autobiographical memories in and outside the laboratory: how different are they from voluntary autobiographical memories?' *Memory and Cognition*, vol.36, no.5, pp.920–32.

Schuman, H., Belli, R.F., and Bischoping, K. (1997) 'The generational basis of historical knowledge', in Jodelet, D., Pennebaker, J., and Paez, D. (eds.) *Political Events and Collective Memories*, London, Routledge.

Sehulster, J.R. (1996) 'In my era: evidence for the perception of a special period in the past', *Memory*, vol.4, pp.145–58.

Shalev, A.T., Peri, T., Canetti, L., and Schreiber, S. (1996) 'Predictors of PTSD in injured trauma survivors: a prospective study', *American Journal of Psychiatry*, vol.153, no.2, pp.219–25.

Sheldon, K.M. and Elliot, A.J. (1999) 'Goal striving, need satisfaction, and longitudinal well-being: the self-concordance model', *Journal of Personality and Social Psychology*, vol.76, no.3, pp.482–97.

Singer, J.A. and Salovey, P. (1993) *The Remembered Self*, New York, The Free Press.

Stadthagen-Gonzalez, H. and Davis, C.J. (2006) 'The Bristol Norms for age of acquisition, imageability and familiarity', *Behavior Research Methods*, vol.68, pp.598–605.

Strauman, T.J. (1996) 'Stability within the self: a longitudinal study of the structural implications of self-discrepancy theory', *Journal of Personality and Social Psychology*, vol.71, no.6, pp.1142–53.

Thorne, A. (1995) 'Developmental truths in memories of childhood and adolescence', *Journal of Personality*, vol.63, no.2, pp.138–63.

Tulving, E. (1972) 'Episodic and semantic memory', in Tulving, E. and Donaldson, W. (eds.) *Organization of Memory*, New York, Academic Press.

Tulving, E. (1983) *Elements of Episodic Memory*, Oxford, Clarendon Press.

Tulving, E. (2002) 'Episodic memory: from mind to brain', *Annual Review of Psychology*, vol.53, pp.1–25.

Tulving, E. and Thompson, D.M. (1973) 'Encoding specificity and retrieval process in episodic memory', *Psychological Review*, vol.80, pp.352–73.

van der Kolk, B.A., van der Hart, O., and Marmar, C.S. (1996) 'Dissociation and information processing in posttraumatic stress disorder', in van der Kolk, B.A., McFarlane, A.C., and Weisaeth, L. (eds.) *Traumatic Stress*, New York, Guilford Press.

Woike, B. (1995) 'Most-memorable experiences: evidence for a link between implicit and explicit motives and social cognitive processes in everyday life', *Journal of Personality and Social Psychology*, vol.68, no.6, pp.1081–91.

Woike, B., Gershkovich, I., Piorkowski, R., and Polo, M. (1999) 'The role of motives in the content and structure of autobiographical memory', *Journal of Personality and Social Psychology*, vol.76, no.4, pp.600–12.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

CHAPTER 18

CONSCIOUSNESS

Jackie Andrade

1 INTRODUCTION

Consciousness is probably the most fascinating and challenging subject of psychological research. Although we know that much of human cognition occurs at a subconscious level, most of us feel that it is our conscious thoughts that form our personalities and inspire our actions. Over centuries, philosophers have provided frameworks for investigating consciousness, possible solutions to some of the problems of consciousness, and vocabularies for discussing the human mind. Only more recently have psychologists begun to research consciousness in its own right. Much of this recent research concerns the biological aspects of consciousness. It uses neuroscience techniques, such as recording the electrical activity of individual neurons, to discover how brain activity differs when we are conscious, or conscious of something, from when we are unconscious, or unconscious of something. However, progress is also being made with cognitive approaches to consciousness. Cognitive psychology has helped define the functions of consciousness, telling us how conscious processes differ from unconscious processes and suggesting possible evolutionary functions of consciousness. These different approaches mean that consciousness research is still a frontier area of psychology, in which explorers still lack agreement about how to explain consciousness or even how to define the problem, yet are already generating many exciting results.

This chapter aims to explain briefly the historical and philosophical roots of consciousness research, and then to discuss the place of consciousness as a concept in contemporary cognitive psychology. It then considers empirical studies of aspects of consciousness and cognitive accounts of consciousness.

It is difficult to give a coherent account of this topic because, although many areas of cognitive psychology inform our understanding of consciousness, these areas are not well integrated and few researchers are trying to pull together the different findings to create a grand theory of consciousness. Much recent cognitive psychology research in the field of consciousness studies has focused on unconscious cognition (the terms unconscious, non-conscious, and implicit cognition are used interchangeably in much of the literature). Baars (1988) recommends contrasting conscious and unconscious cognition and using the differences between them to infer the functions of consciousness. He calls this procedure ‘contrastive analysis’. Many studies do not do this, however, but instead just focus on demonstrating truly unconscious learning or memory. We will look at some of these studies to see whether there is convincing evidence that we can remember or learn without being aware of doing so.

Many cognitive psychologists research high level processes that are apparently dependent on consciousness, for example visual attention, **working memory**, and mental imagery. As an example of this research, we will look at a relatively old study by Schneider and Shiffrin (1977) that helps show the conditions under which automatic and controlled processes operate. Automatic processes are relatively unconscious, in the sense that we have little awareness of their operation, whereas controlled processes are associated with conscious awareness of what is being processed. The concept of controlled processing appears in working memory (see Chapter 9), where the central executive controls the operation of the phonological loop and visuo-spatial sketchpad, and in models of action selection, where controlled processing enables us to behave in novel ways rather than acting through habit. We shall discuss working memory and a model of action selection at the end of this chapter when we consider different ways of explaining consciousness.

Dissociations between conscious and unconscious processes might lead to suggestions that we have a specific module, or modules, for consciousness. We will consider how cognitive neuropsychology, the study of the effects of brain injury, can shed light on the issue of where if anywhere consciousness occurs. The main thrust of the chapter, though, is exploring the functions of consciousness. We will look briefly at how studies of altered states of consciousness can complement more conventional studies of cognition in suggesting hypotheses about the functions of consciousness. I will argue that although consciousness appears to be associated with particular cognitive processes, for example selective attention, all we really know is that these processes are correlated with conscious awareness of stimuli in the environment or in memory. They are the **cognitive correlates of consciousness**. Discovering these correlates of consciousness does not explain conscious experience – why it feels the way it does to see blue or remember a face or imagine a voice – but it does help us to understand the possible role of consciousness in cognition. This role seems to include integrating selected information from different processing modules and making that information available across the cognitive system so that it can guide our behaviour.

The term 'consciousness' means different things to different people, so we begin by trying to define what it is we want to study.

1.1 Defining consciousness

Before reading on, try Activity 18.1.

What was on your list of features of consciousness? For many people, one of the most salient aspects of

✱ ACTIVITY 18.1

Before reading further, spend five minutes thinking about what it means to be conscious. Make a list of the special features of consciousness.

COMMENT

Consciousness can be thought of in different ways. There is the state of consciousness, in the sense of being awake and aware of ourselves in our environment, rather than being asleep and more or less oblivious to what is happening around us. Also under this heading come altered states of consciousness brought about by drugs or hypnosis. Then there is consciousness in the sense of awareness of particular sensations or mental events. Thus while reading this sentence you may be conscious of someone entering the room or of a craving for coffee, but unconscious of the hardness of your chair or the hum of distant traffic. This sense of consciousness has been termed 'access consciousness'. 'Phenomenal consciousness' refers to the particular qualities of our conscious experiences; what it feels like to taste beer or hear a flute, for example. Finally there is self-consciousness, our awareness and monitoring of what we are doing, feeling, thinking, and so on.

being conscious is that we have a feeling of control over our behaviour and even over our thoughts. We feel that we act in a particular way because we decided to act that way, that we have free will. Although it may be difficult to stick to our new year's resolutions, we can modify our behaviour in less ambitious ways. If you get a headache while reading this, you may decide to stop reading – and behave accordingly. Alternatively, you may choose to ignore your headache and attend to the chapter because you wish to finish reading it before going out. Resolving, deciding, choosing, ignoring, feeling pain, attending, and wishing are all aspects of our conscious mental life.

Our conscious experience seems fairly continuous. William James (1890) described it as a 'stream of consciousness'. We have a coherent and persistent awareness of ourselves and of our environment, and are unaware of brief or inconsequential changes in our sensory input. For example, you were probably unaware of this page disappearing from view last time you blinked.

Consciousness is not only about our ability to control our behaviour or to know what is going on around us. It is also about feelings and experiences, for example the smell of spices as you walk past a restaurant, the taste of chocolate, the sensation of jumping into a cold swimming pool or relaxing in a hot bath, the particular feel of looking at something red (rather than something green). Philosophers use the term 'qualia' to describe these qualitative, subjective, experiential aspects of consciousness.

Consciousness has so many different features that, before studying it, we need to know whether it is actually a single thing or several quite different things inappropriately called by the same name. Could a theory of consciousness in principle explain all the different aspects of consciousness that we have discussed so far, or will we need different solutions to different problems of consciousness? This chapter will say little about states of consciousness, being awake rather than asleep for example, although it briefly discusses what altered states of consciousness could reveal about the cognitive correlates of consciousness. It also says nothing about self-consciousness. Rather, it focuses on *consciousness of* particular stimuli or mental events; that is, awareness of particular sights, sounds, memories, ideas, mental images, and so on.

Even if we limit our discussion to consciousness *of* things, there are still two aspects of this type of consciousness to consider. There is the consciousness itself, being aware rather than unaware of something, and there is the experience that this consciousness engenders, what it feels like to taste chocolate or perceive green, for instance. Block (1995) argues that we should treat these two aspects of consciousness as separate problems. He uses the term *phenomenal consciousness* for the experiential aspects of the problem and *access consciousness* for the problem of how, when we are conscious of something, we are able to name it, remember it, decide whether to pick it up, etc. The term 'access consciousness' captures the idea that the contents of consciousness are accessible to other cognitive processes; thus we can talk about our memories or remember things we said. Access consciousness describes this cross-talk between different cognitive modules (a module is a set of mental processes acting together and separately from other sets of processes). Block argues that cognitive psychology only addresses the problem of access consciousness, despite sometimes claiming to solve the problem of phenomenal consciousness as well.

Chalmers (1996) makes a similar distinction. He refers to the problem of how information is shared between modular neural and cognitive systems as the 'easy problem' of consciousness. Empirical research into vision, memory, attention, decision making, and so on addresses the easy problem. The 'hard problem' of consciousness, according to Chalmers, is to explain how and why the neural or cognitive processes of vision, memory, etc. give us the phenomenal conscious experiences of seeing colours or enjoying happy recollections. In a similar vein, Levine (1983) argued that there is an 'explanatory gap' between understanding the neural or cognitive basis of consciousness and explaining the phenomenology. There seems to be nothing about neural or cognitive processes that necessitates their being accompanied by particular experiences. Even if we knew everything about the structure and function of the visual system, could that ever be sufficient to explain why it feels the way it does to see red? This chapter discusses the extent to which cognitive psychology has helped advance our understanding of access consciousness, but while you are reading it, keep asking yourself whether a full understanding of access consciousness would be sufficient to explain phenomenal consciousness.

1.2 Philosophical approaches to consciousness

This section gives a very brief introduction to philosophy of mind, so called because many of the issues pertain to mental processes and states in general, and not merely to conscious processes and states. This section is not intended to be a tutorial on philosophy of mind, just an overview of some of the philosophical issues facing researchers wanting to explain consciousness (in both the access consciousness and phenomenal consciousness senses).

Let's look at the problem of the explanatory gap more closely. Perhaps the reason it is so difficult to relate cognitive and neural brain processes to conscious experience is that they are two entirely different things. Here are three examples of the ways in which they appear to differ:

1. Phenomenal quality. Imagine looking at a particularly bright, warm shade of red. How can interactions between neurons or modules in your brain be 'bright' or 'warm' in the way your experience is? If you imagine a hot cup of black coffee, presumably nothing in your brain turns hot or black.
2. Intentionality. Philosophers describe conscious states such as desiring, believing, and perceiving as 'intentional', meaning they are about things. You can't just desire, you have to desire *something*. It is hard to see how brain states can be about things in the way that mental states are.
3. Spatial position. Neurons are physical entities so they take up space. One neuron can be to the left or right of another, but it does not make sense to talk about mental entities such as images or beliefs having spatial positions.

1.2.1 Mind–body dualism

Dualists such as Descartes solve this mind–body problem by arguing that the mind and the brain are entirely different things. The mind consists of an immaterial 'mindstuff' whereas the brain, like the rest of the body, is made of matter – water, protein, lipids, etc. For consciousness researchers, this is a defeatist stance because it means that the mind does not obey natural scientific laws and is not amenable to scientific investigation. There are other strong objections to dualism. Perhaps the most important is that it does not explain how the mind interacts with the brain or body. How can a thought about drinking water make our physical hand move to pick up a glass and take a sip unless the thought is also somehow physical?

1.2.2 Monism

The converse of dualism is monism, the idea that mind and body are essentially the same thing. Philosophers and scientists who assume that consciousness is a property of the physical brain are called materialists. How do materialists deal with the explanatory gap? One way is to take an extreme view known as eliminative materialism. Proponents of this view, such as Patricia and Paul Churchland, argue that the apparent explanatory gap arises because we use mentalistic terms such as 'desire' and 'belief' that have no scientific basis. We should eliminate these terms from our scientific vocabulary and concentrate on investigating the underlying neuroscience of consciousness. They compare our current use of mentalistic terms with the use of the term 'phlogiston' (once thought to be a substance that escapes when matter burns), which was abandoned when new theories of natural science explained the role of oxygen in combustion. Most materialists do use mentalistic terms but argue that conscious states are brain states and concentrate on investigating their material basis – the chemical and neuronal interactions that underpin consciousness.

Functionalists, the vast majority of whom are also materialists, take a different approach to researching consciousness. Functionalism views mental states as functional or causal states, defined by the ways in which they transform some input (an external stimulus or the product of an earlier cognitive process) into output (information passed to another cognitive module or an overt behaviour). Conscious states are not just **epiphenomena** – by-products of brain processes that have no effect in themselves. Rather, they are the direct causes of our behaviour. Functionalists use the analogy of a computer: the brain is analogous to the hardware of the computer (the silicon chips, wires etc.) and the mind is analogous to the computer's software. The mind is implemented in ('running in', to use computer jargon) the physical brain, in the way that word-processing software might be implemented in a personal computer. The mind could also be implemented in some other physical system, such as a computer, just as a particular software package could run on different sorts of computers. A logical extension of this position leads us to 'strong artificial intelligence' (strong AI), the argument that, if we could program a computer with the same 'software' as a human, then it would be conscious in the same way as us. A less extreme position, weak AI, assumes that computers can have similar 'mental' properties to humans but that there might be something special about biological entities (e.g. carbon-based sensory systems) that make us conscious in the particular way that we are.

Functionalism lies at the heart of cognitive psychology. It means that cognitive psychologists can focus on investigating mental functions without much reference to the brain biology that underpins them. Thus cognitive approaches to consciousness focus on explaining the mental processes that cause one conscious state or another, rather than investigating physical brain activity during that conscious state. It is partly a question of finding an appropriate level of explanation for the phenomenon. Just as your success in an exam might best be explained in terms of your level of attention during lectures, the amount of revision time devoted to your notes, and so on, rather than in terms of biological memory processes such as long-term potentiation, so might consciousness best be explained in terms of cognitive processes.

1.3 The place of consciousness within cognitive psychology

This section provides a brief reminder of the history of cognitive psychology, to help explain why cognitive psychologists are sometimes ambivalent towards the topic of consciousness, with some using it as a variable in their research but few studying it directly. Early in the history of experimental psychology, Wundt trained 'observers' to use introspection (to 'look into' their minds) to give detailed reports on their mental and emotional responses to stimuli. Introspectionism foundered partly because of the subjective nature of the data it produced. When two observers disagreed, it was not possible for an objective third person to resolve their disagreement by looking into their minds and deciding who was reporting their mental states more accurately, or indeed whether their mental states were the same or different. Other problems for introspectionism included research into unconscious, and hence unreportable, processes (e.g. Freud's writings on the contribution of unconscious urges to adult behaviour and Helmholtz's investigations of unconscious processes in vision) and a new emphasis on the functions rather than the structure of mental processes. For example, James (1890) suggested that the function of short-term memory was to keep in consciousness events that have just passed.

These changes paved the way for a radical shift in the way human behaviour was studied. Behaviourists argued that psychologists should concern themselves with objective data, i.e. publicly observable behaviour, rather than subjective introspections. Although some behaviourists included mental events in their explanations of behaviour, the emphasis was on studying the relationships between external stimuli and overt, behavioural responses. Consciousness was no longer a respectable topic for psychological research.

✱ ACTIVITY 18.2

Think of cognitive theories from other chapters. What role does consciousness play in these theories? Do any of the theories help to explain consciousness?

COMMENT

Although many cognitive theories include concepts like attention or working memory, they generally do not specify what processes or qualities make us conscious of some stimuli or cognitive products. For example, are we conscious of information by dint of it being in short-term memory, as James suggested? More recent researchers have argued that we are only fully aware of a subset of representations in working memory (e.g. McElree, 2001). The relationship between consciousness and working memory is discussed in Section 4 of this chapter.

Cognitive psychology developed gradually from the middle of the twentieth century onwards, stimulated in part by a wartime need to explain the role of human factors in tasks such as radar monitoring. It aimed explicitly to explain behaviour in terms of mental activity and thus represented a major shift from behaviourism in its attitude towards the mind. Early cognitive theories included components that related to consciousness, such as attention, but they did not tackle the problem of consciousness directly and did not refer to conscious experience. One reason for this shyness of the topic may have been the need to be perceived as rigorously objective and scientific in an era still overshadowed by behaviourism.

Today, consciousness is increasingly considered as a variable in cognition, particularly in learning and memory research where conscious or explicit processing is contrasted with unconscious or implicit processing. Some examples are given in Chapter 8, as well as in Section 2 of this chapter. In the closely related field of neuropsychology, consciousness is also considered in explanations of conditions such as blindsight. However, there is disagreement about whether the issue is really being tackled. Marcel (1988) argued that 'reference to consciousness in psychological science is demanded, legitimate, and necessary'. It is demanded because it is what makes our mental life interesting, what seems to make us who we are. It is legitimate because the concept of consciousness is no less coherent than other concepts in psychology such as intelligence or personality. It is necessary because we are often implicitly studying consciousness even if we profess to be more interested in some other aspect of cognition. For example, if we ask participants simply to press a button when a light flashes, we are still measuring their conscious experience. If they are not aware of the light flashing, they generally won't respond (this does not mean, however, that their conscious experience *caused* their response). Despite Marcel's call for more explicit discussion of consciousness in cognitive psychology, Banks (1993) argued that psychologists are tiptoeing around the issue of consciousness rather as one might tiptoe around to avoid 'waking the insane attic-bound Aunt of a Gothic novel' (p.257). We might mention conscious processes such as attention, mental imagery, or explicit memory, but we do not try to explain consciousness itself (see also Activity 18.2).

Despite behaviourism, introspection did not die out completely as a tool for psychological research. For example, 'think aloud' protocols have been used to study memory rehearsal and problem solving. However, introspection is becoming more widely used. Some of the studies discussed in the next section rely on participants' reports of whether they were aware of experimental stimuli. Note that the current use of introspection usually only assumes that people have insight into the products of their cognitive processes, not that they can report the processes themselves.

SUMMARY OF SECTION 1

- The term consciousness encompasses the state of being awake, our ability to control our behaviour and be aware of our surroundings, and our mental experiences or 'qualia'.
- There is an explanatory gap between understanding the neural and cognitive functions of the brain and explaining conscious experience.
- Cognitive psychologists view mental states as causal states that affect our behaviour.

2 EMPIRICAL RESEARCH: COGNITIVE STUDIES OF CONSCIOUSNESS

This section focuses on three areas of cognitive psychology. Each area tackles the problem of consciousness in part by investigating unconscious processes. Although this may seem perverse, it helps us work out what processes are associated only with consciousness and not with unconscious processing. The first part covers implicit cognition (specifically, implicit memory and learning, where there is no awareness of what is remembered or learned). Implicit cognition research is important because it can help us to define consciousness better by contrasting it with unconscious processes. It raises the question of what, given the extent of unconscious processing, might be the function of consciousness. The second part revisits earlier research into automatic and controlled processing. Although not phrased in terms of unconscious and conscious cognition, these studies show us the essential characteristics of conscious processes. They are slow but flexible whereas automatic or unconscious processes are fast and efficient but inflexible. The third part considers the condition known as blindsight. Neuropsychological studies help elucidate the function of 'normal' consciousness and raise questions about the functional and physical structure of consciousness. This account necessarily misses much of the research in cognitive psychology that relates to consciousness. The areas it does cover are those in which researchers have particularly used their findings to frame questions about consciousness, although even in these areas much research is reported with scant if any mention of what it tells us about consciousness.

2.1 Implicit cognition

2.1.1 Implicit memory

Implicit memory is memory without any accompanying sensation of remembering. It is revealed by changes in performance on specially designed memory tests. For example, if I show you the word 'witness' in the context of an apparently unrelated, non-memory task and then give you a surprise memory test, you may not recall seeing 'witness' or recognize it as a word from the earlier task. However, if I ask you to say the first word that comes to mind starting 'wit—', your implicit memory for the word will make you more likely to say 'witness' than if you had not just seen that word.

Tests such as this *word stem completion task* are often referred to as indirect memory tests because they measure memory without directly asking people to decide if they remember the stimuli. Indirect tests are assumed to measure predominantly implicit memory, whereas direct tests measure explicit memory. Note, though, that no memory test is 'process pure'; performance on almost any memory test can be influenced by both implicit and explicit memory. For example, if you are asked to think of the first word that comes to mind beginning with 'wit–' and nothing comes to mind, you may try to think back to the earlier task to search for clues, for words that might fit the stem. If you remember that 'witness' was one of the words on the first task, and use that as your response, then you are using your explicit memory and the task is not giving a pure measure of your implicit memory.

Two studies of implicit memory are described below. Further examples are given in Chapter 8, but these two are chosen because they appear to show implicit memory in the absence of explicit memory (with the caveat about the process-impurity of memory tests).

In an early study of implicit memory, Eich (1984) showed that prior presentation of a word in a particular context could bias its subsequent interpretation. Participants in Eich's experiment heard a list of word pairs. One word in each pair was a homophone, that is, it sounded like another word with a different meaning and spelling, for example PANE (as in 'pane of glass') is a homophone of PAIN (as in 'stomach pain'). The other word in each pair made clear the intended interpretation of the homophone. 'Window-PANE' and 'taxi-FARE' are examples of the word pairs used by Eich. Note that the homophones he used were the less common interpretations, PAIN and FAIR being the more frequently encountered spellings and meanings. Although participants in Eich's study could hear the word pairs, they could not attend to them because their main task was to shadow (repeat) an essay played at the same time. Memory for the homophones was tested in two ways. On the recognition test, participants listened to a list of words and were asked to say whether each word was old (i.e. present in the unattended list of word pairs) or new. This is a direct test of memory because it requires participants to make a judgement about their memory; it is therefore assumed to measure mainly conscious or explicit memory. Participants were unable to recognize the unattended words. For the second memory test, the experimenter read out a list of words and participants were asked to spell them. Their spelling was biased by the previous presentation of the homophones (e.g. they were biased towards spelling P-A-N-E rather than the more common P-A-I-N). This spelling test is an indirect test of memory because it does not require deliberate recollection. These findings therefore suggest that participants had implicit memory for the homophones, but not explicit memory.

Another example of implicit memory in the absence of explicit recollection is the false fame effect. Jacoby *et al.* (1989) used a task on which participants had to say whether names belonged to famous people. In the study phase of Experiment 2 of their investigation, participants read aloud a list of 40 non-famous names. In the test phase, 10 of these names were mixed with 10 new non-famous names for a recognition memory test. The remaining 30 names were mixed with 30 new non-famous names and 60 famous names for the fame judgement task. Before the fame judgement task, participants were told that the names they had just read were the names of non-famous people, hence if they recognized any name from the first phase of the experiment they should respond 'non-famous'. The experimenters manipulated the degree of attention paid to the names in the study phase. In the full attention condition, participants were told that the experimenters were interested in their ability to pronounce the names quickly and accurately. In the divided attention condition, they were told to pay as little attention as possible to the pronunciation task, concentrating instead on listening to a stream of spoken digits and spotting runs of three odd numbers. The fame judgement task was sufficiently difficult that participants made a number of false positives, saying that a name was famous when in fact it was not. If participants had encountered a non-famous name earlier in the experiment, they were more likely to judge incorrectly that it was famous if the study phase took place under conditions of divided attention. Divided attention impaired explicit recognition of the names and thus reduced participants' ability to use explicit memory to interpret feelings of familiarity. In the absence of conscious recognition, familiar names were assumed to be famous.

One of the most exciting things about this field of research was the discovery that people with amnesia often performed as well as people with normal memory on the indirect tests of memory. In other words, despite their severely impaired explicit memory, amnesics had almost normal implicit memory. For example, Squire and McKee (1992) replicated the false fame effect in amnesic participants. Amnesics were significantly impaired at recognizing the previously presented names compared with the control subjects, but they were just as biased towards judging presented non-famous names as famous. Such findings led to new rehabilitation strategies because they showed that amnesics had the potential to learn new information even though they appeared to have no memory. We'll see an example of a new learning strategy that has been used in rehabilitation later in this chapter (in Box 18.3 on errorless learning). The dissociation between the effects of brain injury on implicit and explicit memory raises questions about the structure of consciousness. We will return to these questions in Section 2.3.

Implicit memory phenomena are often referred to as **priming**. Priming is the improvement in performance caused by previous exposure to the target stimulus or by previous or concurrent exposure to a closely related stimulus. For example, Meyer and Schvaneveldt (1971) showed that participants decided more quickly that pairs of letter strings were two real words when they were related words (e.g. doctor and nurse) than when they were unrelated words (e.g. doctor and cabbage). Building on this study, Marcel (1983) showed that participants identified 'doctor' as a real word faster when it was preceded by a very brief presentation of the word 'nurse' than when it was preceded by an unrelated word. The word 'nurse' in this example is called the prime. Marcel's study is mentioned again below, and is also discussed in Chapter 2. The idea behind priming is that: a) activation of an item's representation in memory lingers, so that the representation is still slightly activated next time the item is encountered, making it easier to reactivate even if only a partial cue is presented, such as a word-stem; b) activation spreads to related representations, making related representations easier to activate than unrelated, unprimed representations. Priming is also used in a more general sense, to refer to the activation of moods or stereotypes (see Box 18.1 on unconscious influences on behaviour). Demonstrations of priming

BOX 18.1 Unconscious influences on behaviour

Research in the field of social cognition suggests that priming may considerably influence our behaviour outside the laboratory. Primes may influence our mood and behaviour without us being aware of them. For example, Bargh *et al.* (1996) asked participants to arrange lists of words to form meaningful sentences. In the experimental group, each word list contained a word related to the concept of old age, for example 'wrinkled' or 'ancient'. Participants were surreptitiously timed as they left the laboratory after completing this task. Those in the experimental group, who had been exposed to the elderly primes, left the laboratory more slowly than those in the control group, who were not exposed to those primes.

Bargh *et al.* argued that the primes activated a stereotype of old age and participants behaved in accordance with that stereotype even though they had not noticed the primes. Neumann and Strack (2000) showed that people's mood can be affected by the mood of others around them, even when they are unaware of the cause of their mood change. When participants listened to text read in a sad voice, they were more likely to rate their own mood as sad but were unaware that their mood had changed as a result of listening to the sad voice. Lieberman (2000) argues that implicit cognitive processes, such as priming of stereotypes and mood states, underlie the phenomenon commonly known as 'intuition', that is, our ability to judge social situations and respond appropriately without being aware of the information (other people's moods, etc.) on which we base our judgements.

show that we are not aware of all our knowledge, or of the basis of our behaviour. Priming is often preserved even when brain injury causes impairments to explicit cognition; thus people with amnesia typically have preserved implicit memory despite their impaired explicit memory. Box 18.3 on errorless learning in Section 3 mentions a way of using this preserved priming in rehabilitation. Despite preserved implicit memory, lack of explicit memory has an impact on the more general aspect of consciousness that James (1890) termed our 'stream of consciousness'. Baddeley (1990) cites the example of Clive Wearing, whose amnesia was so severe that he repeatedly noted in his diary that he had just regained consciousness. Memory and consciousness thus appear to be correlated, but the direction of causation is unclear: does normal consciousness require intact memory function or does normal memory function require consciousness? We will return to this problem of correlational evidence at the end of the chapter.

2.1.2 Implicit learning

Studies of implicit memory show that we can 'remember' things without having any conscious experience of remembering them. A more profound claim has been made by researchers in the field of implicit learning, namely that we can learn things without ever being aware of them. If this claim is true, it helps establish some ground rules for our study of consciousness by telling us what is possible without consciousness. Evidence that a lot of learning is possible without consciousness might suggest that consciousness is just an epiphenomenon that plays no causal role in our cognition. We will therefore look closely at some of the evidence for implicit learning and at some of the methodological problems that face researchers trying to show that participants had no awareness of the material they learned.

One way of demonstrating this unconscious or implicit learning is to present the stimuli to be learned very quickly, too quickly for participants to notice more than just a flash on the screen. This is called subliminal presentation. The study by Marcel, mentioned above, has become a controversial classic: a classic because it demonstrated priming even though the presentation of the primes was apparently subliminal, and controversial because of claims that the findings could not be replicated when stricter definitions of 'subliminal' were used. However, subsequent researchers have used subliminal presentation with other test procedures to provide further evidence for implicit learning.

Some of this evidence comes from demonstrations of the mere exposure effect, the tendency for people to prefer stimuli they have encountered before even if they were unaware of them during the previous encounter. For example, Kunst-Wilson and Zajonc (1980) presented novel black and white patterns subliminally. Even though participants said they could not see the patterns because they were presented so briefly, they later tended to select those patterns when presented with pairs of patterns (one presented earlier and one new) and asked to choose the one they preferred. On a recognition test in which participants chose the pattern they remembered seeing earlier from each pair, participants performed at chance, ie. they were just guessing. So their first, unconscious, encounter with the patterns apparently changed their emotional response to them even though they had no conscious or explicit memory of seeing the patterns before (see also Chapter 16).

As suggested by the controversy over Marcel's (1983) study, there are problems with using subliminal presentation to demonstrate unconscious learning. One problem is ensuring that all stimuli are subliminal for all participants. This is tricky because some stimuli are easier to perceive than others. For example, you can sometimes hear someone say your name even if you don't hear anything else they say because you are attending to another conversation (Moray, 1959). Another limitation is the equipment used to present the stimuli. Early studies used tachistoscopes, boxes that were specially designed to show stimuli for very brief and accurately timed periods. Many researchers now use computers for running their experiments, but computerized presentation times are limited by factors like the screen refresh rate – how quickly the computer redraws the display. A refresh rate of 17 msec means that stimuli can only be presented for multiples of 17 msec. This limitation

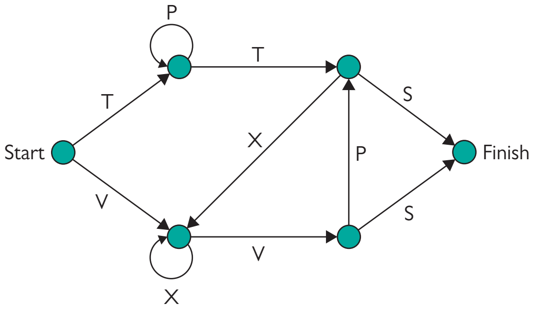


FIGURE 18.1 The artificial grammar used by Reber (1967). Grammatical letter strings are created by following the arrows through the array from left to right. The curved arrows indicate items that can be repeated.

makes it hard for the experimenter to present a stimulus for long enough to effect some learning but briefly enough to prevent the participant identifying the stimulus.

A solution to these problems is to present stimuli supraliminally (for long enough that they can be consciously perceived), but to test learning of some hidden relationship between them. For example, Reber has claimed that people can implicitly learn hidden rules, constituting an 'artificial grammar', that underpin a set of supraliminally presented letter strings (see Activity 18.3). Although they cannot verbalize their knowledge, it allows them to distinguish grammatical from ungrammatical items with above-chance accuracy. Two of Reber's early experiments are discussed below. An example of an artificial grammar is shown in Figure 18.1.

Reber (1967) asked participants to learn 28 letter strings or 'sentences' as part of a 'memory experiment'. The sentences were presented in seven sets of four sentences. Each sentence was viewed for 5 seconds, after which the participant tried to write it. After attempting to reproduce the four sentences in a set, the participant was told which sentences they had reproduced correctly and the procedure was repeated until they reached the criterion of reproducing all four sentences correctly on two consecutive trials. The next set of sentences was then presented. For participants in the experimental group, the sentences were formed according to the rules of an artificial grammar. For those in the control group, the sentences contained randomly ordered letters. The results were as follows: both groups learned the second of the seven sets of sentences in fewer attempts than the first. However, from the third set onwards, the control group continued to make a mean of between 8 and 11 errors per set whereas the experimental group continued to improve, making a mean of only 3 errors on the seventh set. Reber argued that the experimental group acquired knowledge of the grammar that enabled them to learn more efficiently.

A second experiment in Reber's study showed that prior exposure to the grammatical sentences enabled participants to distinguish new grammatical sentences from ungrammatical sentences. Participants learned 20 sentences in a procedure similar to that described above. They were then tested on 88 trials with new sentences, comprising two presentations each of 22 grammatical sentences (that had not been encountered in the learning phase) and 22 ungrammatical sentences. Participants were told that the sentences they had already learned were grammatical, and were

✱ ACTIVITY 18.3

First, trace through the network shown in Figure 18.1 to convince yourself that TPPPTS and VXVPS are grammatical strings whereas VTPS and TPTTTS are ungrammatical. Now try writing down or explaining to a friend the rules of this grammar.

COMMENT

Verbal report is often used as a way of finding out whether participants in implicit learning experiments were aware of the information they learned. The stimuli used in these experiments are typically complex and novel and therefore difficult to describe. You may have struggled to report the rules of Reber's grammar even though you were looking at Figure 18.1 while doing so. The problem of eliciting participants' explicit knowledge through verbal report is addressed later in this section.

asked to use their knowledge of those sentences to decide if each test sentence was grammatical or ungrammatical. Their decisions were correct on a mean of 79 per cent of trials, well above the mean of 50 per cent expected from chance.

Subsequent studies of artificial grammar learning have used procedures similar to that of this second experiment. Participants learn a set of grammatical strings to a predetermined criterion, and then attempt to distinguish novel grammatical strings from ungrammatical strings. Performance is typically above chance, though not as impressive as in Reber's study, even though participants cannot state the rules they used to decide which were the grammatical items. This apparently implicit learning of the grammar is dissociable from explicit recall of the grammatical strings presented in the learning phase. For example, Knowlton *et al.* (1992) found that amnesic patients were as good as controls at classifying novel strings as grammatical or ungrammatical, but were poorer at recognizing exemplars that had been encountered in the learning phase.

Nissen and Bullemer (1987) used an alternative procedure for testing implicit learning of hidden regularities between visible stimuli. They gave control and amnesic participants a choice reaction time task in which they had to watch a panel of four lights (ABCD) and, whenever a light came on, to press the key under that light as quickly as possible. Participants were not told that the lights came on in a fixed order (the 10-item sequence DBCACBDCBA was presented repeatedly). Controls and amnesics got faster at this task until the sequence was switched to a random order; at this point, their reaction times increased. The amnesic participants showed no awareness that the lights had come on in a regular sequence. Control participants were also often unaware of the sequence, particularly if they performed the key pressing task under conditions of divided attention.

A problem with these demonstrations of implicit learning is that we have no way of determining participants' awareness of the key stimuli or relationships while they are doing the task. If we ask them if they are aware of the grammar, for example, then we draw their attention to it and lose the opportunity for demonstrating learning without awareness. So researchers have to ask participants afterwards what they were aware of during the task. This is unsatisfactory, because it relies on people's memory of what they were aware of rather than measuring awareness on-line. Another problem is that quite a small amount of knowledge may be enough to boost people's performance above the chance level. For instance, they may not know the whole grammar, just that a certain letter can be repeated or come at the start of a letter string. Knowing possible starts for grammatical strings may be sufficient to distinguish a few of the grammatical strings from the ungrammatical strings, resulting in performance that is slightly above baseline. If a test of awareness simply asks 'Were you aware of the grammar?' or 'What was the grammar?' then it will miss the knowledge that actually boosted performance on the grammar test, and that knowledge may well be explicit. Participants may interpret the question as asking for a complete report of the grammar, which they cannot give, and so do not volunteer their knowledge of fragments of the grammar. These problems are discussed by Shanks and St John (1994). They argue that experimenters must use tests of explicit knowledge, or awareness, that meet two criteria before they can claim that learning resulted in truly implicit knowledge. The information criterion states that the test of awareness must probe for the sort of information that could support performance on the test of learning (for example, knowledge that a particular letter often comes at the start of a grammatical letter string). The sensitivity criterion states that the test of awareness must be sensitive to all the relevant explicit knowledge; it must be just as sensitive as the test of implicit knowledge. Simply asking participants to state the rules of the artificial grammar fails on both counts. It does not prompt them to report fragments of the grammar and it does not give them any recall cues, whereas the grammar judgement test presents the actual grammatical stimuli and these may serve as cues to memory.

Given the difficulty of ensuring lack of awareness of critical stimuli in awake participants, perhaps a better strategy would be to study learning in people who are unconscious. Testing patients receiving anaesthetics offers a way of tackling this issue, though one that presents more difficulties than might at first be imagined. One difficulty is that depth of anaesthesia, or 'degree of unconsciousness', fluctuates during an operation and there is not yet a universally agreed way of measuring this fluctuation or of establishing exactly the depth of anaesthesia at which a person loses consciousness, in the sense of losing all awareness of themselves and their surroundings. Thus a finding that patients can learn information presented during anaesthesia may reflect explicit learning during undetected moments of consciousness, rather than truly unconscious learning. Another difficulty is that the sensitivity of the memory tests has not been established so, if a study shows no evidence for learning during anaesthesia, this may be due to use of a test that is too insensitive to detect small amounts of preserved learning. Not surprisingly, although many studies have investigated learning during anaesthesia, their findings have been mixed (Andrade, 1995).

Catherine Deeprose and I obtained evidence for priming during anaesthesia in a study that overcame some of the problems discussed above (Deeprose *et al.*, 2004, 2005). We played words (e.g. tractor) to patients during surgery. When the patients came round from the anaesthetic, we asked them to respond to word stems (e.g. tra–) with the first word that came to mind. Playing a word during surgery increased the likelihood of patients using that word to complete a word stem on recovery. In other words, they showed some implicit memory for the words even though they were anaesthetized while receiving them. We had pilot tested our word stem completion test to ensure that it was reasonably sensitive and also reasonably uncontaminated by explicit memory (it was relatively unaffected by a manipulation of attention known to affect explicit memory). Thus we gave ourselves a good chance of demonstrating implicit memory for words played during surgery. We minimized the chance of priming occurring during moments of awareness by using an EEG measure of depth of anaesthesia throughout word presentation and testing patients who were unparalysed, because the drugs that are often used to paralyse patients during surgery make it even harder to detect moments of consciousness. We are therefore reasonably confident that we have demonstrated that memories can be primed in someone who is unconscious. The next step is to show that new information can be learned during anaesthesia.

To summarize, there is some evidence for learning without consciousness of what is learned. However, this implicit learning is rather difficult to demonstrate convincingly and is also not very useful, in the sense that we cannot revise, contemplate, or tell people about what we have learned implicitly. We cannot select what we learn when learning implicitly, and we cannot retrieve the learned material voluntarily. It would appear that conscious processes (e.g. the active selection, rehearsal, and elaboration of information) contribute to much of our everyday learning. Even so, implicit learning may help us to pick up repeated patterns or relationships among stimuli or events, and by doing so help us direct our conscious learning processes towards interesting features of our environment. Implicit memory is easier to demonstrate. When only a small amount of learning has occurred, because of inattention or brain damage for example, the resulting memory may be implicit. We may be unaware of what we have learned because the encoded material does not reach some threshold for consciousness or because it has not been processed by a 'conscious memory module'. This issue of the structure of consciousness is discussed briefly in Section 2.3 on the neuropsychology of consciousness.

2.2 Controlled versus automatic processing

The concept of controlled processing is closely allied to that of conscious processing. As you have seen in the previous section, if we want to demonstrate implicit learning or memory, we have to make it very difficult for participants to process the target material in an active way, for example by distracting their attention from it. This sort of active processing is often known as controlled processing (as opposed to automatic processing). The idea of controlled processing is central to concepts such as working memory, discussed in Section 4 as a potential model of consciousness.

Controlled cognitive processes typically accompany consciousness, that is, they are cognitive correlates of consciousness. This section therefore describes a classic study by Schneider and Shiffrin (1977) that defined and demonstrated controlled and automatic processes in visual attention.

Schneider and Shiffrin (1977) based their theorizing on Atkinson and Shiffrin's (1968) model of memory (see Chapters 8 and 9), arguing that automatic processes operate on the long-term memory store (an interconnected array of nodes) whereas controlled processes require the limited capacity short-term store, essentially the currently activated nodes of the long-term store. They defined automatic processes as the activation of a sequence of nodes in the long-term store via connections between those nodes that have become relatively permanent through repeated use. Once triggered, automatic processes operate without active control so it is difficult to stop them or change their course. In contrast, activation of a novel sequence of nodes requires attention, which limits our capacity to activating just one novel sequence at a time but gives us control over the activation.

Schneider and Shiffrin demonstrated the difference between these two processing modes using 'target search tasks' that required participants to detect targets among arrays of distractors as quickly and accurately as possible. They manipulated: a) the number of targets participants had to search for; b) the number of items (targets and distractors combined) on each slide or 'frame'; and c) the mapping between the set of targets for any series of trials (the 'memory set') and the distractor set. This mapping manipulation was the key to demonstrating automatic and controlled processing modes. In the *consistent mapping* condition, the targets were always selected from the same set of items and the distractors were always selected from a different set so, for example, participants might search for target digits among distractor letters. In the *varied mapping* condition, the targets and distractors were drawn from the same set, so participants might search for letters among letters and a particular letter could be a target on one trial and a distractor on another. Figure 18.2 gives examples of trials in these different conditions.

The consistent and varied mapping conditions produced quite different patterns of response times. In

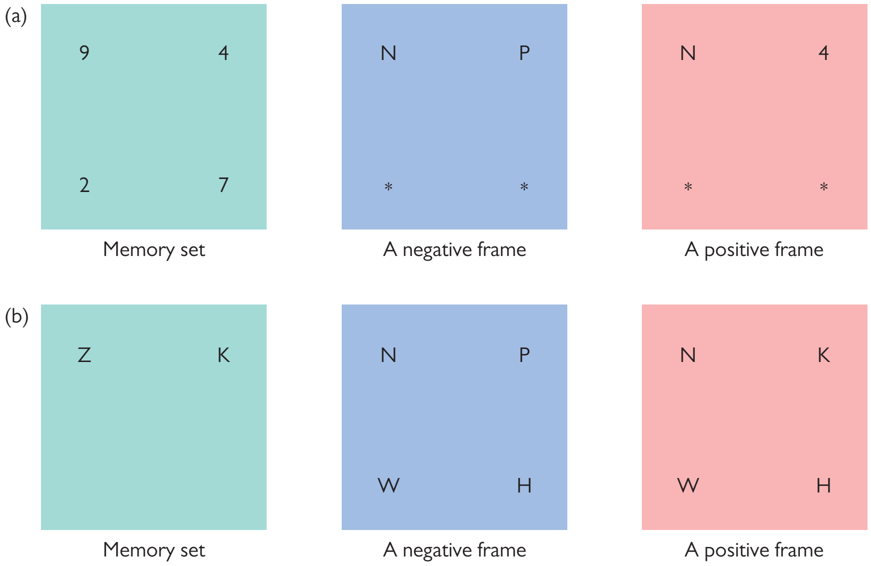


FIGURE 18.2 Examples of stimuli from Schneider and Shiffrin's (1977) experiments. On a given trial, participants saw and memorized the set of targets for that trial (the 'memory set'). They then saw a sequence of frames. Their task was to detect whether any of the memorized targets appeared in the sequence. (a) Looking for digits among letters: stimuli from a consistent mapping condition with a memory set size of 4 and a frame size of 2. The asterisks represent pattern masks in the positions not occupied by targets or distractors. (b) Looking for letters among letters: stimuli from a varied mapping condition with a memory set size of 2 and a frame size of 4.

the consistent mapping condition, performance was fast and relatively unaffected by memory set size and frame size. In the varied mapping condition, performance was slower and was slowed further still by increasing the number of targets to be searched for and the number of items per frame to be searched. Schneider and Shiffrin argued that the varied mapping condition necessitated a serial search through the array in each frame, hence the larger the array, the longer the search. In contrast, the consistent mapping condition allowed the items in the arrays to be searched in parallel because all we have to do is spot a digit (say) among letters, and we have had many years of practice at recognizing digits. Because we only have to spot a digit, there is no need to maintain the specific identities of the memory set items in short-term memory.

Subsequent researchers have extended Schneider and Shiffrin's description of controlled processes to cover a variety of cognitive processes that are flexible but slow and expensive in terms of cognitive resources. In everyday terms, we use controlled processing when performing novel tasks or wanting to override habitual behaviours. For example, when making black coffee for a friend, we may have to attend to each step of the procedure to stop ourselves adding milk. If talking to our friend at the same time, it is easy to lapse into automatic behaviour and make the 'action slip' of adding milk to their coffee as well as to ours. Automatic processes are typically fast and efficient but inflexible. We are generally unaware of the operation of automatic processes. For example, an experienced tennis player will hit a ball without stopping to think how to do so whereas a novice may have to think about how to hold the racquet, how hard to hit, and so on. Automatic processes make little impact on explicit memory, so someone driving a car on 'autopilot' may arrive at their destination safely, but with little memory of the journey. Thus, the concepts of automatic and controlled processing map closely onto the currently more fashionable concepts of unconscious or implicit processing and conscious or explicit processing. But this mapping raises another question: is consciousness something we use to control our behaviour, or do we *become* conscious of our behaviour when we exert control over it?

2.3 The neuropsychology of consciousness

The research presented in the previous sections raises questions about the structure of consciousness. Are we conscious of a stimulus (or of a memory) because it exceeds some threshold of salience or activation, or are we conscious of it because it is processed in a particular way? In other words, is consciousness something that might be associated with very many cognitive processes or is it a feature of particular cognitive modules? Is there even a 'consciousness module'? There are no clear answers to these questions at present, but this section aims to show the potential for neuropsychology to help find answers. While you are reading it, bear in mind that the search for a consciousness module may be futile: Dennett (1991) argues that finding a place where consciousness happens – a '**Cartesian theatre**' – will do nothing to help us explain consciousness. We will meet Dennett again in Section 4.

Studies of altered consciousness following localized brain injury provide a way of assessing whether consciousness is a unitary or modular function. If it is a unitary function, it could be localized in a single 'consciousness area' of the brain, or distributed across a network of interconnected brain regions, or it could be the result of some non-localized process such as synchronized activity across brain regions (see ffytche, 2000). If consciousness is a modular function, then our conscious awareness of colours might be caused by processes localized in quite different brain regions from, say, our conscious awareness of movements or sounds.

At first glance, neuropsychological studies suggest that consciousness is modular because brain injury causes loss of consciousness of only a subset of sensations and cognitions. For example, people with amnesia are not conscious of information learned since the onset of their amnesia. People with unilateral neglect, on the other hand, have normal consciousness of their memories but lack consciousness of one side of space (see Chapter 13). Is it the case then that our consciousness for memories is separate from our consciousness for space, as a modular interpretation would suggest? The answer is not clear. Patients with amnesia may lack consciousness of their memories because they lack critical unconscious processes that feed into a unitary consciousness. Likewise neglect patients may not have suffered damage to a 'consciousness of space' module but rather have deficits in attentional processes that feed into a unitary consciousness.

Zeki and ffytche (1998) studied a blindsight patient known as GY. Blindsight is a disorder exhibited by some patients with brain damage leading to blindness in part of the visual field. In blindsight, there is a somewhat preserved ability to respond appropriately to visual stimuli in the blind region of the visual field, despite having no sense of seeing them. GY's blindness is selective, so that he reports having some sort of conscious experience of some visual stimuli in his blind field (e.g. fast-moving stimuli) but denies having any experience of other stimuli (e.g. slow-moving stimuli). Usually, GY can identify a stimulus from a small selection of distractors only if he has some conscious experience of it. Occasionally, however, he can identify stimuli even without any conscious experience, that is, he exhibits blindsight for these stimuli. Zeki and ffytche (1998) used functional magnetic resonance imaging (fMRI) to compare GY's brain activity with and without conscious perception of visual stimuli. GY could detect the direction of movement of slow- and fast-moving stimuli, but he was usually only conscious of the fast stimuli. The two types of stimuli differentially triggered activity in motion cortex, the increase in activity being greater with the fast-moving stimuli for which GY reported some conscious experience. Thus consciousness of visual stimuli appears to be related to the amount of activity in a localized brain area, a brain area specialized for processing that type of stimulus rather than a general 'consciousness centre'. ffytche (2000) suggests that the increased localized activity associated with consciousness may reflect the activity of additional populations of neurons in that brain region or it may reflect more complex processing by neurons that are also active when we are not conscious of the stimulus to which they respond. Note, however, that although Zeki and ffytche's data support the hypothesis of a modular consciousness, our overall conscious experience – our individual 'stream of consciousness' – may still reflect the aggregation of processing across many brain regions. Consciousness for motion may be dissociable from consciousness for colour, for instance, but these two consciousness modules must somehow be bound together to produce our normal conscious experience of a moving coloured stimulus.

SUMMARY OF SECTION 2

- Studies of priming or implicit memory show that people retain more information than they are aware of remembering.
- Studies of learning without awareness of what is learned have been hotly debated; so-called implicit learning might reflect failure to detect small amounts of awareness. A recent study of learning during anaesthesia suggests that memory priming occurs even when patients are unconscious.
- Schneider and Shiffrin (1977) argued that controlled search processes operate serially and are slow, increasingly so as task demands increase, because they depend on limited capacity short-term memory systems. Automatic search processes are fast and can operate in parallel because they operate on well-learned pathways in long-term memory. Automatic processing typically happens without awareness, whereas controlled processing is associated with conscious awareness of the task in hand and explicit memory for its products.
- Neuropsychological studies can help determine the structure of consciousness. ffytche (2000) used a study of blindsight to argue that consciousness is a function of modular brain systems. Alternative arguments are that is a single 'consciousness module' or that consciousness is a distributed function.

3 WHAT IS CONSCIOUSNESS FOR?

Try Activity 18.4 before you read on.

The studies of implicit learning and memory discussed in the previous section suggest that it is possible to learn about a variety of different stimuli without being conscious of them. We appear to 'remember' things without any conscious experience of doing so. If we can do this without consciousness, does consciousness actually serve any function or is it an epiphenomenon, a by-product of brain processes that does not in itself affect the system? This section discusses evidence that consciousness serves a variety of functions that together make us able to function effectively even in novel environments.

Research into automatic and controlled processes shows that automatic or unconscious processes tend to be fast and efficient but inflexible. Providing there is no competition for sensory systems (e.g. trying to view two complex pictures simultaneously) or response effectors (e.g. trying to write two answers simultaneously), several automatic processes can run concurrently. Controlled or conscious processes, on the other hand, are slower and more demanding of cognitive resources so it is hard to carry out more than one at a time. Stimuli must therefore be selected for conscious processing. The Cheshire cat activity, discussed later in Activity 18.5, Section 3.1, illustrates this point. Chapter 2 on attention discusses how this selection is done. Despite their disadvantages, we need conscious processes when performing new tasks or trying to override habits. This is illustrated by the tendency of people to generate stereotyped responses when distracted from a random generation task, a task that requires frequent strategy shifts to avoid lapsing into stereotyped response patterns. Baddeley *et al.* (1998) asked participants to generate random sequences of digits or key presses. They challenged the availability of controlled processing resources by asking participants to perform other tasks at the same time (for example, solving problems or retrieving information from long-term memory). In these conditions, faster and more automatic processes generated stereotyped responses such as '1, 2, 3' or parts of familiar telephone numbers. Conscious or controlled processes thus seem to be associated with flexible responding. Box 18.2 on affective priming suggests that conscious processes help us make rational rather than emotional decisions. The research presented in Box 18.3 on errorless learning shows that making mistakes prevents people with amnesia from learning as effectively as they might. With normal memory, consciousness of our memory for past errors may help us adapt our behaviour by learning from our mistakes.

✱ ACTIVITY 18.4

Make a list of the features and functions of consciousness revealed by the research discussed in the previous section.

COMMENT

The studies of automatic and controlled processing suggest that automated, relatively unconscious processing is fast and efficient, helping us to deal with large amounts of information. However, automatic processes make little impact on memory and the studies of implicit cognition suggest that memory and learning without awareness are of limited use in directing our behaviour. They may serve to bias our behaviour – to give us a 'hunch' about the correct answer – but they do not give us the confidence to alter our behaviour deliberately as a result of what we have learned. For example, people often don't particularly enjoy doing artificial grammar learning tests, because they feel they are just guessing and do not have strategies for maximizing their performance on the test.

BOX 18.2 RESEARCH STUDY Affective priming

Zajonc (1980) hypothesized that emotional or affective responses can be triggered by information that undergoes only minimal processing (see also Chapter 16). Cognitive responses require more processing. This is known as the 'affective primacy hypothesis'. Murphy and Zajonc (1993) tested this hypothesis in a series of experiments that compared the influence of subliminal and 'optimal' (i.e. consciously visible) primes on responses to subsequent stimuli.

In the first experiment in their study, Murphy and Zajonc asked 32 participants to rate their liking of Chinese ideographs, on a scale of 1 = did not like the ideograph at all, to 5 = liked the ideograph quite a bit. Each ideograph was shown for 2 seconds. There were four types of trial:

- *No prime controls*, where the ideographs were shown alone.
- *Irrelevant prime controls*, where a geometric shape preceded each ideograph.
- *Positive prime trials*, where a photograph of a happy face preceded each ideograph.
- *Negative prime trials*, where a photograph of an angry face preceded each ideograph.

For participants in the subliminal prime condition, each slide of a shape or face was presented for just 4 msec, followed immediately by an ideograph that served both as the stimulus to be rated and as a visual mask to prevent the image of the prime lingering in iconic memory. Participants in the optimal prime condition viewed each prime slide for 1 second, followed immediately by the ideograph. A summary of these experimental trials is shown in Figure 18.3.

The results are shown in Table 18.1. The ideographs were rated as more pleasant when preceded by a subliminal positive prime, compared with the no prime and irrelevant prime control conditions, and more unpleasant when preceded by a subliminal negative prime. In contrast, the visible or 'optimal' primes had no effect on ratings of the ideographs.

In a follow-up experiment, Murphy and Zajonc tested the effects of subliminal and optimal primes on ratings of the size of object represented by each ideograph. This time the relevant primes were small or large shapes and the irrelevant primes were faces with neutral expressions. The results contrasted with those of the experiment described above: the subliminal primes had no effect on ratings of size whereas the large optimal primes led to higher ratings and the small optimal primes to lower ratings.

Conclusion

Murphy and Zajonc (1993) argued that the affective primes altered participants' mood, in the sense that participants had an emotional response to the primes. When participants were aware of their mood change and of its source, they could ignore it when rating the ideographs. However, when they

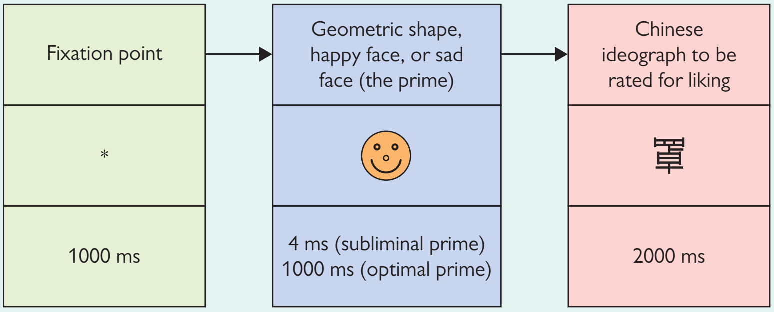

Fixation point
*
1000 ms
Geometric shape, happy face, or sad face (the prime)
4 ms (subliminal prime)
1000 ms (optimal prime)
Chinese ideograph to be rated for liking
罩
2000 ms

FIGURE 18.3 Experimental trials used to demonstrate affective priming.

TABLE 18.1 Mean ratings of liking for Chinese ideographs in affective prime, irrelevant prime, and no prime conditions

Condition	Subliminal presentation (4 msec)	Optimal presentation (1000 msec)
Positive prime	3.46	3.02
Negative prime	2.70	3.28
Irrelevant prime	3.06	3.15
No prime	3.06	3.11

were unaware of it, their affective response to the primes 'spilled over' onto the rating task. For negative primes, the authors referred to this effect as 'free floating anxiety' – that is, anxiety without awareness of what caused it or what we are anxious about. Consciousness of the primes allowed participants to override their affective response to them when judging the ideographs.

BOX 18.3 RESEARCH STUDY Errorless learning

Baddeley and Wilson (1994) argued that one of the main functions of explicit memory is to help us learn from our mistakes. Without awareness of our errors, past mistakes serve only to prime similar mistakes in the future. They tested their hypothesis by comparing two modes of learning in 32 participants with normal memory (16 young and 16 elderly adults) and 16 participants with amnesia who were assumed to have normal implicit memory combined with impaired explicit memory. Participants with normal memory received one list of ten words in the errorful learning condition and another list of ten words in the errorless learning condition. Amnesic participants received five words in each condition, to avoid floor and ceiling effects. The stimulus words were all five letters long and chosen because their two-letter stems could be completed in several ways. For example, the stem QU- could be completed as QUOTE (the stimulus in this case) or QUIET, QUEEN, QUACK, etc. The large number of potential completions maximized the possibility for making errors in the errorful learning condition.

In this errorful learning condition, participants were told that the experimenter was thinking of a 5-letter word beginning QU- and were asked to guess what the word might be. After making up to four incorrect guesses, the participant was told that the word was QUOTE (or a back-up word if they happened to guess the target straight away) and asked to write it down. This procedure was repeated for the other words in the list, and then again two more times for the entire list.

In the errorless learning condition, participants were told 'I am thinking of a five-letter word beginning with QU and that word is QUOTE; please write that down' (p.55). The list of target words was presented three times, as in the errorful condition.

These first three trials were termed the pre-training phase. Learning condition was only manipulated during this pre-training phase. The test phase comprised nine further learning trials. On each of these trials, regardless of the initial learning condition, the experimenter provided the first two letters of each word in the list and asked the participant to write down a word starting with those letters from the earlier list. If they could not remember a word from the previous phase, they were asked to say any word that came to mind beginning with those letters. In the case of incorrect responses, the experimenter provided the correct word. Baddeley and Wilson analysed performance in the test phase in terms of the probability of learning, that is, the probability of an item that is not known on one trial becoming learned on the next trial. The learning probabilities for the two learning conditions are illustrated in Figure 18.4, which shows that the participants with amnesia benefited considerably more from the errorless learning procedure than the young and elderly participants with normal memory. Without explicit memory for their errors, amnesic participants were unable to correct their mistakes on subsequent trials and so found the learning task particularly difficult when they were encouraged to make errors on the initial, pre-training trials. Errorless learning techniques have been used to teach amnesics useful information such as how to program a personal organizer to remind them of appointments.

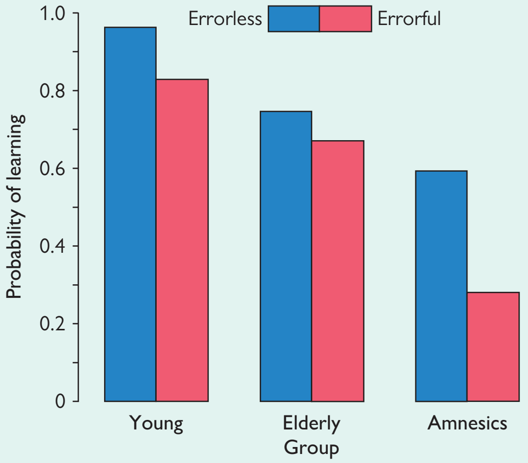


FIGURE 18.4 Learning as a function of learning condition. *Source: Reprinted from Baddeley, A. and Wilson, B.A. (1994) 'When implicit learning fails: amnesia and the problem of error elimination', Neuropsychologia, vol.32, no.1, pp.53–68, with permission from the publisher, Pergamon*

An important feature of consciousness is that it seems to break the modularity of mind. Cognitive psychologists often assume that cognitive processes are modular, that is, that they operate in clusters that function independently from other clusters of processes. Each cluster or 'module' processes a particular sort of information, so there may be one module for comprehending spoken language and another module for recognizing faces. Consciousness involves cross-talk between these otherwise independent modules. If we are conscious of something, we can talk about it, decide to touch it or ignore it, imagine holding or owning it, and have beliefs about it. In studies of blindsight, note that patients do not tend to initiate responses towards stimuli. Even though they can respond appropriately to stimuli when encouraged to guess, their residual processing of the stimuli is of little practical use to them. Without the conscious experience of seeing, someone with this condition would not, for example, pick up a glass of water placed in their blind field even if they were thirsty and could point to its location when encouraged to do so. Normally, seeing a glass of water means that we can also drink it or talk about it – consciousness of the visual percept makes it available to action and language modules. In blindsight, although some visual information is processed and influences 'guessing' behaviour, that information is not sufficient to break out of the visual perception module to form a basis for conscious behaviour. Likewise with implicit learning. Although exposure to covert regularities may improve performance on an implicit learning task with similar stimuli, even without awareness of what has been learned, there is limited transfer of this improvement to tasks with different stimuli constructed according to the same rules (e.g. Gomez, 1997). When we are aware of what we have learned, we are better able to apply that learning to novel tasks.

Although cognitive research has helped identify the functions of consciousness in controlling our behaviour and allowing cross-talk between cognitive modules, it does not explain why these functions are associated with conscious experience. Could it be possible for a computer, say, to perform these functions without consciousness? Philosophers use the term 'zombie' to refer to the idea of someone exactly like us, with the same cognitive processes as us, the same knowledge, memories, planning abilities, and so on, but without consciousness. If this idea is conceivable, if it makes sense to us even though zombies are a fiction, then it suggests that the 'hard problem' of consciousness is indeed a very hard problem because there is nothing about our cognitive processes that *necessitates* the conscious experience that accompanies them (see Chalmers, 1996).

3.1 Altered states of consciousness

The cognitive basis of altered states of consciousness is not well understood. Nonetheless, altered states are interesting because, by providing a contrast, they help us

BOX 18.4 Detecting consciousness: a real-life problem

Does consciousness do anything? Think carefully before you answer this question (and look back at Activity 18.4) because the answer matters for the real-life problem of detecting consciousness in other people, not to mention other animals, computers, and robots. If consciousness does something, then we should be able to find a difference between the behaviour of organisms that have consciousness and those that do not. If consciousness is an epiphenomenon, an inessential luxury of feeling that accompanies our thoughts and behaviours but does not influence them, then we can only truly know about our own consciousness (see Nagel's (1974) classic: 'What is it like to be a bat?').

Searle (1980) explores the problem of detecting consciousness with his **Chinese Room** thought experiment. In this thought experiment, Searle is in a room receiving questions in Chinese about a Chinese story and sending out answers in Chinese. He does this by following a set of rules that tell him how the questions map onto the story script and how to translate the script into answers in Chinese. Searle speaks English but not Chinese. To an observer on the outside of the Chinese Room, his behaviour is the same as if he were doing the task in English, but for Searle inside the room it is very different. In English, he has understanding; in Chinese he does not. Searle uses the Chinese Room to argue that computers will never be conscious – all they can do is simulate consciousness, as he simulates understanding Chinese in the Chinese Room. In relation to detecting consciousness, the important point is that we, as outside observers, cannot tell the difference.

The premise of the **Turing test**, proposed by the mathematician Alan Turing as a test of computer intelligence, is that we can tell if a computer is intelligent, or conscious, by comparing its responses to those of an intelligent and conscious human. If they are indistinguishable, then we should conclude that the computer is conscious like the human. This test assumes that consciousness makes a difference to our answers; there could not be zombies. It also assumes that computer consciousness will be like human consciousness. If what it is like to be a computer is very different from what it is like to be a human, then conscious computers will fail the Turing test.

How could we detect consciousness if someone or something is unable to communicate? This is an everyday problem for anaesthetists, because the amount of anaesthetic needed to make someone unconscious varies across individuals and operations. It is difficult to judge accurately a patient's level of unconsciousness, or 'depth of anaesthesia', using physiological signs such as heart rate. This problem is being solved by new depth of anaesthesia monitors that provide on-line analysis of the brain's electrical activity (EEG) during surgery. However, these EEG measures only tell us the probability that a patient is unconscious. Detecting consciousness requires communication with the patient, for example by asking them to answer questions by moving their hand. If patients are paralysed during surgery, this can be done using the so-called 'isolated forearm technique' where a tourniquet is used to prevent paralysing drugs reaching the patient's lower forearm, allowing them to move their hand in response to questions should they regain consciousness during the operation (see Andrade *et al.*, 2008). This method is used much less than EEG measures.

People in vegetative states pose similar problems. New research suggests that brain imaging techniques could help inform arguments about continuing life support treatment for unresponsive patients. Owen and colleagues (2006) reported the case of a patient with a diagnosis of vegetative state who, when asked to imagine hitting a tennis ball or moving through her house, showed brain activity in exactly the same areas of the brain that support mental imagery in conscious subjects. She later recovered consciousness.

Derbyshire (2006) argues that consciousness requires representational memory. A fetus or newborn infant cannot feel pain, for example, because he or she has not yet learned that that sensation is pain. Learning what pain is – acquiring a 'pain representation' – requires interactions with the world and other humans. Consider the ethical ramifications of this position, and indeed of any argument that only humans have sufficiently large and complex brains to support consciousness. At what point in development does an individual's brain become sufficiently large and complex? Vohs and Schooler (2008) provided a neat, and ethically innocuous, demonstration that how we answer such questions – what we believe about consciousness – affects how we behave. Participants in their study who read arguments against free will were more likely to cheat on a maths test than those who read neutral statements.

reflect on what 'normal' consciousness is like. Because they are generally dysfunctional states, in the sense that they are not conducive to normal everyday behaviour, they give us clues about the functions of normal conscious states. Altered states are therefore included briefly in this chapter as a discussion point to help you think about the possible functions of consciousness.

Drugs such as ketamine and lysergic acid diethylamide (LSD) cause hallucinations and other perceptual disturbances such as synaesthesia, a condition where stimuli in one sensory modality trigger experiences in another sensory modality (e.g. touching something hard may produce the sensation of seeing green). It appears that these drugs cause a flooding of the sensory system and a breaking down of the modularity of sensory systems. Normal consciousness may therefore involve selecting incoming sensory information to prevent too much information reaching higher level cognitive processes. Drugs such as alcohol cause loss of inhibition, making us more likely to say things or do things that we would refrain from doing in our normal conscious state. Normal consciousness may therefore involve monitoring and controlling our behaviour.

Hypnosis is a state of deep relaxation that makes people more susceptible to suggestion. There is debate about whether it is truly an altered state of consciousness, or merely a response to the particular combination of relaxation and social pressure to conform to the hypnotist's suggestions. Nonetheless, hypnotized people can perform surprising feats, such as speaking 'forgotten' languages from their childhood or even undergoing minor surgery without painkillers or anaesthetic. One explanation of such feats is that hypnosis reduces our normal tendency to check our mental contents against the outside world. This checking and updating of our mental model of ourselves and our environment is called reality monitoring. With reduced reality monitoring, hypnotized people become more credulous because they are less likely to check the hypnotist's suggestions against what they know to be true. This makes them better able to maintain a 'hallucination' even if it contradicts incoming sensory information.

Altered states of consciousness suggest that normal consciousness operates with or on a selected portion of the information that constantly bombards our senses. For a demonstration of this selection, try the Cheshire Cat activity described in Activity 18.5. Normal consciousness also involves checking our current mental state against incoming information from our environment and checking our behaviour against our intended goals.

✱ ACTIVITY 18.5

The Cheshire cat illusion (Duensing and Miller, 1979)

Sit facing a picture of a cat (or a real one if you can persuade it to sit still) on an otherwise blank wall, with another blank wall to your right. Hold the edge of a mirror against your nose, and tilt it so you view the cat picture with your left eye only. Hold up your right hand so that it is reflected in the mirror, the reflection being viewed by your right eye. Move your right hand slowly towards and away from you. With appropriate adjustments to the mirror or to the direction of hand waving, you should experience the Cheshire cat illusion: moving your hand appears to 'rub out' parts or all of the cat.

When you experience the Cheshire cat illusion, your brain is receiving two separate visual images, yet you only see one of them at a time. What does this tell you about consciousness?

Note: if you have difficulty seeing this illusion, try this simpler version. Roll a sheet of paper into a tube and look through it with your left eye while viewing your right hand with your right eye. Hold your right hand about the same distance from your eye as is the end of the tube, then move your hand slightly to and fro until you perceive a hole in it with the same diameter as the tube.

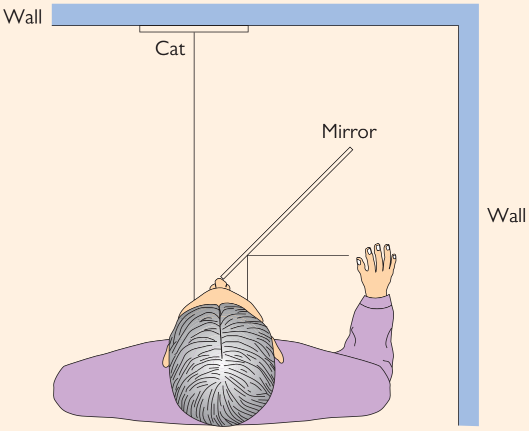


FIGURE 18.5 Set-up for the Cheshire cat illusion.

COMMENT

The illusion occurs because the two percepts – the one of the cat and the one of your hand – cannot be fused. Only one percept can be conscious at a time, hence if you are conscious of seeing the reflection of your hand, you will not be conscious of seeing the part of the cat picture that occupies the same apparent visual location as the reflection of your hand. This illusion illustrates the selective and limited capacity aspect of consciousness. The phenomenon is known as 'binocular rivalry' and has been used by neuroscientists to determine the neuronal activity associated with consciousness of a visual stimulus (e.g. Logothetis and Schall, 1989).

SUMMARY OF SECTION 3

- Consciousness is associated with feelings of behavioural control. It helps us override habitual and emotional responses and to learn from our mistakes.
- If consciousness plays a causal role in our behaviour or thought processes, then we should be able to detect or measure consciousness in others by observing their behaviour or changes in brain activity.
- The effects of mind-altering drugs and of hypnosis suggest that normal conscious states involve monitoring and control of behaviour, and selection of incoming information. We are only conscious of a small amount of information at any time.

4 COGNITIVE THEORIES OF CONSCIOUSNESS

This section focuses on three theories: working memory, global workspace theory, and higher order thought theory. Baddeley's working memory model is chosen because it is already a widely used model in cognitive psychology (see Chapter 9). If it can also say something about consciousness then it helps us integrate our thinking about consciousness with our existing understanding of other aspects of cognition. Baars' global workspace approach is discussed as a contrast, partly because it is the most detailed and comprehensive cognitive model of consciousness currently available and partly because Baars explicitly discusses the hypothetical relationship between consciousness and working memory. Finally, higher order thought theory (actually a class of theories) is included because it gives a central role to mental representation and has led to interesting speculations about why consciousness evolved.

Short-term memory, and later working memory, have often been identified with consciousness. Thus James (1890/1918) described memory generally as a way of bringing back past conscious experiences, but argued that for primary memory there is nothing to bring back because 'it was never lost; its date was never cut off in consciousness from that of the immediately present moment' (1918, p.647). Baddeley (1993) portrays working memory as a conduit to consciousness that serves to bring together information in different modalities from perception and long-term memory, enabling us to imagine novel solutions to problems of evolutionary significance. For example, he suggests that a vivid image of a hunting ground, which included locations where food was found before and locations where danger lurked, would have evolutionary significance as a tool for predicting events and planning action. In a similar vein, my colleagues and I have argued that mental imagery, which is heavily dependent on working memory resources, serves as a 'motivational bridge' to keep goals in mind while we work towards achieving them (Elaborated Intrusion theory; Kavanagh *et al.*, 2005).

By combining information from different modalities into conscious representations or images, working memory presents one solution to the binding problem. This is the problem of creating a coherent, unified conscious experience from disparate sensory inputs (note, though, that the binding problem also pertains at lower levels of cognition, so we also need to bind together the results of location and object detection processes, for example). Baddeley (2000) added a new component to the working memory model, called the episodic buffer, as a temporary store for bound information and as an interface between working memory, long-term memory, and consciousness. However, the revised model does not specify whether information is conscious by virtue of being stored in the episodic buffer or whether it is conscious only when acted upon by central executive processes. Nonetheless, the wealth of evidence for a role of working memory in conscious cognition, such as problem solving and mental imagery, suggests that working memory is a cognitive correlate of consciousness.

The hypothetical central executive of working memory seems particularly closely correlated with consciousness, playing a key role in conscious, strategic activities such as retrieval of information from long-term memory and selective attention (Baddeley, 1996). Baddeley (1986) used Norman and Shallice's (1986) cognitive model of contention scheduling and supervisory attentional system (SAS) as a model of the central executive. Norman and Shallice's model aimed to explain action selection. They hypothesized that routine responses are selected by a relatively automatic process called contention scheduling, essentially the selection of habitual action schemata by virtue of their high activation level and ability to inhibit competing schemata. Changes in behaviour are effected by a SAS system that increases the activation of a non-habitual action schema so that it 'wins' in the contention scheduling process. The SAS thus serves as an error-correcting mechanism and as the basis for conscious behavioural control, allowing us to respond appropriately in novel situations. Failure of the SAS explains errors in behaviour, such as everyday action slips (e.g. habitually adding milk when intending to make a cup of black coffee) and perseverative behaviours following frontal brain injury (i.e. persisting with an established response when altered conditions mean a new response is needed). Baddeley's adoption of the SAS as a model of the central executive extended its remit from the conscious control of action to the conscious control of cognition generally. However, both models are subject to the criticism of postulating little more than a 'homunculus', a little man in the head that tells us which action to do or which memory to retrieve.

Dennett (1991) refers to the homunculus problem as the problem of the Cartesian theatre: defining a cognitive or neuroanatomical module that 'does' consciousness does not help us explain consciousness, it merely tells us where it occurs. Dennett's **multiple drafts** theory of consciousness offers a solution to this problem. Dennett argues that it is misleading to think of consciousness as something that suddenly happens, in the sense that stimulus processing works its way up from low-level sensory processes to higher-level cognitive processes and somewhere along the way our representation of the stimulus suddenly enters consciousness. The idea that information is processed and then 'sent' to consciousness raises all sorts of awkward questions about what sort of homunculus or 'inner eye' it is sent there for. Dennett's theory aims to get around these awkward questions. Stimuli are not processed and then sent to a consciousness module, they are just processed. Which of many parallel streams of processing we become aware of, and how we experience them, depends on how and when the system is 'probed' by tasks that require particular responses (see Activity 18.6).

Whereas Baddeley argues that working memory is necessary for consciousness, Baars (2002) argues that consciousness is necessary for working memory. All elements of 'active working memory', such as subvocal rehearsal and visual imagery, are conscious. Baars describes consciousness as a 'global workspace', a means of bringing together the products of processing from widely distributed modules. This bringing together – of inner speech, mental imagery, strategic recall, etc. – is necessary for working memory to function. Voluntary control of behaviour is also dependent on consciousness, requiring conscious goals and conscious perception of the effects of our actions. Thus we can learn from our mistakes or avoid emotions biasing our behaviour because our behavioural control processes have access to our knowledge of past mistakes or the causes of our mood swings. Baars (1997) uses the analogy of a theatre. The unconscious processes of syntax analysis, visual boundary analysis, semantic processing, etc. are the stagehands working behind the scenes in the theatre of consciousness. Although there are actors on the stage, we only see the actor currently performing in the spotlight. Working memory forms the stage of consciousness, and representations in

✱ ACTIVITY 18.6

Quickly make a list of what you are conscious of right now.

COMMENT

What was on your list? Did it include the feel of the chair against your legs, how warm or cold you feel, a tickle or an itch? Are you conscious of any of those sensations now? If the answer is 'yes', then ask yourself this: of the items on your original list, were you conscious of any of them *before* you started making the list? If not, what were you conscious of then? This activity shows that, although we feel intimately acquainted with our own conscious experience, that experience changes according to the task we are doing. There is no coherent 'stream of consciousness', just 'multiple drafts' with the potential to become conscious.

working memory are the actors on the stage. Actors only step into the spotlight when chosen by the stage director – likewise the contents of working memory have the potential to become conscious but usually do not do so unless selected by the central executive. Once in the spotlight, the actor can be seen by everyone else: once a representation becomes conscious, it is accessible to other cognitive processes. Thus consciousness may appear to overcome some of the modularity of mind, enabling us to talk about our ideas, express our feelings, use remembered information to solve problems, and so on (though see also Chapter 20).

Baars (2002) draws on many recent neuroscience studies to support his theory. The essence of these findings is that unconscious processing of stimuli activates localized brain regions, whereas conscious processing of the same stimuli activates widely distributed brain regions. For example, an fMRI study by Dehaene *et al.* (2001) showed that processing of masked visual words was associated with activation in early visual cortex, whereas processing of visible, unmasked words was also associated with activation in parietal and prefrontal cortex. However, the finding by Zeki and ffytche (1998) of increased but still localized activation with conscious perception seems to contradict Baars' theory. Thus it appears that consciousness is often but not always associated with global brain activity.

Whereas Baars argues that information is conscious when it is globally accessible, higher order thought (HOT) theories (e.g. Rosenthal, 1997) argue that information is conscious when it is the subject of a thought. Although my senses are constantly bombarded, I only consciously perceive those sensations that lead, via cognitive feedback loops, to the thought that I am having then, 'I am seeing yellow' or 'I'm hungry' (Humphrey, 2006). Pause for a moment to consider how a higher order thought theory might explain your answers to Activity 18.6. Sensing the world around us and also thinking about those sensations requires considerable brain power, so higher order thought theories predict that only organisms with large and complex brains will be conscious. Being able to represent our own sensations and feelings sets us up well to be able to represent other people's feelings, in other words to develop 'theory of mind'. If there is something that it is like to be me, then I can infer that there is something that it is like to be someone else, an essential step towards predicting how other people will behave and therefore a building block of social interaction (Humphrey, 2002).

SUMMARY OF SECTION 4

- Working memory is closely allied to consciousness, providing a way of binding together information from perception and long-term memory to create conscious, multi-modal representations.
- Although working memory is a cognitive correlate of consciousness, the relationship between working memory and consciousness is unclear. Baars (1997) suggested that working memory serves to select the information that will become conscious. More recently, he has argued that consciousness is necessary for working memory (Baars, 2002).
- Baars sees consciousness as a global workspace, enabling cognitive modules to share information. 'Global access' to conscious information is consistent with findings of widespread brain activity when participants are aware of a stimulus, contrasted with more localized activity when they are unaware of it.
- According to higher order thought theories, we are conscious of information that is the subject of a thought. Consciousness requires a large and complex brain and provides us with insight into our own mental states and those of other people.

5 CONCLUSION: WHAT CAN COGNITIVE PSYCHOLOGY TELL US ABOUT CONSCIOUSNESS?

Cognitive psychology has helped to generate hypotheses about the functions of consciousness. For example, studies of errorless learning and of priming have suggested that consciousness helps us control our behaviour by avoiding repeating our mistakes and avoiding emotional rather than rational responses. However, although consciousness is associated with behavioural control, we have no evidence that consciousness *causes* behavioural control. Similarly, we have seen that consciousness is associated with limited resource systems like working memory and selective attention, but we do not know whether consciousness plays a causal role in remembering or attending. Thus cognitive psychology has helped to discover the cognitive correlates of consciousness, that is, those cognitive processes that always accompany consciousness. But any correlation only tells us that two things are related, not that one causes the other. This problem is exemplified by the debate outlined in Section 4 about whether working memory is necessary for consciousness or vice versa. Hardcastle (2000) discusses this problem of correlational data in relation to the search for the neural correlates of consciousness. She argues that there will be many correlates of consciousness, but to explain consciousness effectively we must identify the 'proximal cause'; the correlate that is the most important for consciousness. She uses the example of depression: is Fred depressed because he has received bad news or because the level of noradrenalin in his brain has dropped? If people are more likely to become depressed as the result of neurotransmitter changes than as the result of hearing bad news, then neurotransmitter levels are the more important correlate of depression. Hardcastle suggests that it is too early to be able to say which of the many correlates identified so far is the proximal cause of consciousness. Indeed, it is not even clear whether the answer will lie with those who seek to reduce consciousness to particular brain modules or individual neural events, or with those who argue for a dynamic systems approach, who would suggest that consciousness emerges from complex and global interactions of brain processes. In the meantime, cognitive psychology provides us with useful techniques for analysing and researching consciousness. At least, it is beginning to address the so-called easy problems of consciousness – its role in attention, recollection, voluntary action, and so forth. The problem of why consciousness feels the way it does remains a very hard problem.

FURTHER READING

Blackmore, S. (2010) *Consciousness: An introduction, 2nd edition*, London, UK, Hodder Education. A comprehensive and accessible introduction to philosophical and empirical issues in consciousness studies. Includes interesting practical exercises to help you think about key problems. What I really like about this book is that it begs to be read rather than merely dipped into, providing explanation as well as information about the key issues.

Blackmore, S. (2001) 'Consciousness', *The Psychologist*, vol.14, no.10, pp.522–5. A succinct overview of the most interesting issues at a very basic level – a good starting point.

There are several excellent collections of chapters on consciousness that offer compulsive reading once you start browsing:

Metzinger, T. (ed.) (2000) *Neural Correlates of Consciousness*, Cambridge, MA, MIT Press. Particularly recommended are the chapters by Hardcastle and ffytche.

Velmans, M. and Schneider, S. (eds.) (2007) *The Blackwell Companion to Consciousness*, Oxford, Blackwell. Schneider's chapter, 'Dennett on the nature of consciousness', is a good introduction to Dennett's work.

Block, N., Flanagan, O., and Güzeldere, G. (eds.) (1997) *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA, MIT Press. The introduction by Güven Güzeldere gives a clear and comprehensive overview of philosophical and empirical issues.

Bayne, T., Cleeremans, A., and Wilken, P. (eds.) (2009) *The Oxford Companion to Consciousness*, Oxford, Oxford University Press. An alphabetized series of entries ranging from 'absent qualia' to 'zombies'.

REFERENCES

Andrade, J. (1995) 'Learning during anaesthesia: a review', *British Journal of Psychology*, vol.86, no.4, pp.479–506.

Andrade, J., Deeprose, C., and Barker, I. (2008) 'Incidence of awareness and memory priming in paediatric surgery with general anaesthesia', *British Journal of Anaesthesia*, vol.100, pp.389–96.

Atkinson, R.C. and Shiffrin, R.M. (1968) 'Human memory: a proposed system and control processes', in Spence, K.W. and Spence, J.D. (eds.) *The Psychology of Learning and Motivation, Vol. 2*, New York, Academic Press.

Baars, B.J. (1988) *A Cognitive Theory of Consciousness*, New York, Cambridge University Press.

Baars, B.J. (1997) *In the Theater of Consciousness: The Workspace of the Mind*, New York, Oxford University Press.

Baars, B.J. (2002) 'The conscious access hypothesis: origins and recent evidence', *Trends in Cognitive Sciences*, vol.6, no.1, pp.47–52.

Baddeley, A.D. (1986) *Working Memory*, Oxford, Oxford University Press.

Baddeley, A.D. (1990) *Human Memory: Theory and Practice*, Hove, UK, Lawrence Erlbaum Associates Ltd.

Baddeley, A.D. (1993) 'Working memory and conscious awareness', in Collins, A.F., Gathercole, S.E., Conway, M.A., and Morris, P.E. (eds.) *Theories of Memory*, Hove, UK, Lawrence Erlbaum Associates Ltd.

Baddeley, A.D. (1996) 'Exploring the central executive', *Quarterly Journal of Experimental Psychology*, vol.49, pp.5–28.

Baddeley, A. (2000) 'The episodic buffer: a new component of working memory?' *Trends in Cognitive Science*, vol.4, no.11, pp.417–23.

Baddeley, A.D. and Wilson, B.A. (1994) 'When implicit learning fails: amnesia and the problem of error elimination', *Neuropsychologia*, vol.32, pp.53–68.

Baddeley, A.D., Emslie, H., Kolodny, J., and Duncan, J. (1998) 'Random generation and the executive control of working memory', *Quarterly Journal of Experimental Psychology*, vol.51A, no.4, pp.819–52.

Banks, W.P. (1993) 'Problems in the scientific pursuit of consciousness', *Consciousness and Cognition*, vol.2, no.4, pp.255–63.

Bargh, J.A., Chen, M., and Burrows, L. (1996) 'Automaticity of social behavior: direct effects of trait construct and stereotype activation on action', *Journal of Personality and Social Psychology*, vol.71, pp.230–44.

Block, N. (1995) 'On a confusion of a function of consciousness', *Behavioral and Brain Sciences*, vol.18, no.2, pp.227–87.

Chalmers, D. (1996) *The Conscious Mind*, Oxford, Oxford University Press.

Deeprose, C., Andrade, J., Harrison, D., and Edwards, N.E. (2005) 'Unconscious memory priming during anaesthesia: a replication', *British Journal of Anaesthesia*, vol.94, no.1, pp.57–62.

Deeprose, C., Andrade, J., Varma, S., and Edwards, N. (2004) 'Unconscious learning during surgery with propofol anaesthesia', *British Journal of Anaesthesia*, vol.92, no.2, pp.171–7.

Dehaene, S., Naccache1, L., Cohen, L., Le Bihan, D., Mangin, J.-F., Poline, J.-B., and Rivière1, D. (2001) 'Cerebral mechanisms of word masking and unconscious repetition priming', *Nature Neuroscience*, vol.4, pp.752–8.

Dennett, D. (1991) *Consciousness Explained*, Boston, Little, Brown & Co.

Derbyshire, S.W.G. (2006) 'Can fetuses feel pain?' *British Medical Journal*, vol.332, pp.909–12.

Duensing, S. and Miller, B. (1979) 'The Cheshire Cat effect', *Perception*, vol.8, pp.269–73.

Eich, E. (1984) 'Memory for unattended events: remembering with and without awareness', *Memory & Cognition*, vol.12, pp.105–11.

ffytche, D. (2000) 'Imaging conscious vision', in Metzinger, T. (ed.) *Neural Correlates of Consciousness*, Cambridge, MA, MIT Press.

Gomez, R.L. (1997) 'Transfer and complexity in artificial grammar learning', *Cognitive Psychology*, vol.33, no.2, pp. 154–207.

Hardcastle, V.G. (2000) ‘How to understand the N in NCC’, in Metzinger, T. (ed.) *Neural Correlates of Consciousness*, Cambridge, MA, MIT Press.

Humphrey, N. (2002) *The Mind Made Flesh: Frontiers of Psychology and Evolution*, Oxford, UK, Oxford University Press.

Humphrey, N. (2006) *Seeing Red: A Study in Consciousness*, Cambridge, MA, Harvard University Press.

Jacoby, L.L., Woloshyn, V., and Kelley, C. (1989) ‘Becoming famous without being recognized: unconscious influences of memory produced by dividing attention’, *Journal of Experimental Psychology: General*, vol.118, no.2, pp.115–25.

James, W. (1890/1918) *The Principles of Psychology, Vol. I*, London, Macmillan and Co. Ltd.

Kavanagh, D.J., Andrade, J., and May, J. (2005) ‘Imaginary relish and exquisite torture: the Elaborated Intrusion theory of desire’, *Psychological Review*, vol.112, no.2, pp.446–67.

Knowlton, B.J., Ramus, S.J., and Squire, L.R. (1992) ‘Intact artificial grammar learning in amnesia: dissociation of classification learning and explicit memory for specific instances’, *Psychological Science*, vol.3, no.3, pp.172–9.

Kunst-Wilson, W.R. and Zajonc, R.B. (1980) ‘Affective discrimination of stimuli that cannot be recognized’, *Science*, vol.207, pp.557–8.

Levine, J. (1983) ‘Materialism and qualia: the explanatory gap’, *Pacific Philosophical Quarterly*, vol.64, pp.354–61.

Lieberman, M.D. (2000) ‘Intuition: a social cognitive neuroscience approach’, *Psychological Bulletin*, vol.126, no.1, pp.109–37.

Logothetis, N.K. and Schall, J.D. (1989) ‘Neuronal correlates of subjective visual perception’, *Science*, vol.245, pp.761–3.

Marcel, A.J. (1983) ‘Conscious and unconscious perception: experiments on visual masking and word recognition’, *Cognitive Psychology*, vol.15, pp.197–237.

Marcel, A.J. (1988) ‘Phenomenal experience and functionalism’, in Marcel, A.J. and Bisiach, E. *Consciousness in Contemporary Science*, Oxford, Oxford University Press.

McElree, B. (2001) ‘Working memory and focal attention’, *Journal of Experimental Psychology: Learning, Memory, and Perception*, vol.27, no.3, pp.817–35.

Meyer, D.E. and Schvaneveldt, R.W. (1971) ‘Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations’, *Journal of Experimental Psychology*, vol.90, pp.227–35.

Moray, N. (1959) ‘Attention in dichotic listening: affective cues and the influence of instructions’, *Quarterly Journal of Experimental Psychology*, vol.11, pp.56–60.

Murphy, S.T. and Zajonc, R.B. (1993) ‘Affect, cognition, and awareness: affective priming with optimal and suboptimal stimulus exposures’, *Journal of Personality and Social Psychology*, vol.64, pp.723–39.

Nagel, T. (1974) ‘What is it like to be a bat?’ *The Philosophical Review*, vol.83, pp.435–50.

Neumann, R. and Strack, F. (2000) ‘Mood contagion – the automatic transfer of mood between persons’, *Journal of Personality and Social Psychology*, vol.79, no.2, pp.211–23.

Nissen, M.J. and Bullemer, P. (1987) ‘Attentional requirements of learning: evidence from performance measures’, *Cognitive Psychology*, vol.19, pp.1–32.

Norman, D.A. and Shallice, T. (1986) ‘Attention to action: willed and automatic control of behavior’, in Davidson, R.J., Schwartz, G.E., and Shapiro, D. (eds.) *Consciousness and self-regulation*, New York, Plenum.

Owen, A.M., Coleman, M.R., Boly, M., Davis, M.H., Laureys, S., and Pickard, J.D. (2006) ‘Detecting awareness in the vegetative state’, *Science*, vol.8, p.1402.

Reber, A.S. (1967) ‘Implicit learning of artificial grammars’, *Journal of Verbal Learning and Verbal Behavior*, vol.6, pp.855–63.

Rosenthal, D.E. (1997) ‘A Theory of Consciousness,’ in Block, N., Flanagan, O.J., and Guzeldere, G. (eds) *The Nature of Consciousness*, MIT Press.

Schneider, W. and Shiffrin, R.M. (1977) ‘Controlled and automatic human information processing: 1. Detection, search, and attention’, *Psychological Review*, vol.84, pp.1–66.

Searle, J.R. (1980) ‘Minds, brains, and programs’, *Behavioral and Brain Sciences*, vol.3, pp.417–57.

Shanks, D.R. and St John, M.F. (1994) ‘Characteristics of dissociable human learning systems’, *Behavioural and Brain Sciences*, vol.17, no.3.

Squire, L.R. and Mckee, R. (1992) ‘Influence of prior events on cognitive judgments in amnesia’, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol.18, no.1, pp.106–15.

Vohs, K.D. and Schooler, J.W. (2008) ‘The value of believing in free will: encouraging a belief in determinism increases cheating’, *Psychological Science*, vol.19, pp.49–54.

Zajonc, R.B. (1980) ‘Feeling and thinking: preferences need no inferences’, *American Psychologist*, vol.35, pp.151–75.

Zeki, S. and ffytche, D.H. (1998) ‘The Riddoch syndrome: insights into the neurobiology of conscious vision’, *Brain*, vol.121, pp.25–45.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

CHAPTER 19

COGNITIVE MODELLING AND COGNITIVE ARCHITECTURES

Paul Mulholland and Stuart Watt

1 WHAT IS COGNITIVE MODELLING?

In this chapter we are going to return to a psychological method that was introduced in Chapter 1 and has been raised at various points since: cognitive modelling. You have already met a few cognitive models, although you may not have realized that they were cognitive models at the time. Now is a good time to look at the method of cognitive modelling in a bit more detail and see the contributions that models make to cognitive psychology.

First of all, what is cognitive modelling? **Cognitive modelling** involves building a working model of a cognitive process and then comparing the behaviour of that model against human performance. If the model behaves in the same way as humans, then the structure of the model and the way it works may give some insight into how humans perform the task. Cognitive modelling has been used to help explain a range of cognitive processes such as face recognition (see Chapter 4), language comprehension (Chapter 6), and analogical problem solving (Chapter 10).

Cognitive modelling is very like the kind of technique that car designers or architects use to make it easier to see and test their designs. Psychologists use models in the same way: to make it easier to understand

and to test their theories. Generally speaking, there are two different approaches to cognitive modelling. First, there is a high-level approach. To follow the car analogy, a high-level model should look and behave as much like a car as possible, without necessarily having the same internal workings. A high-level model might state that there is an engine, but might not say exactly how it worked or what it was made of. But there is also a low-level approach, where a modeller would look at representing the kinds of bits that cars are made from (wheels, axles, pistons, valves, and so on) and try to understand how the behaviours of these components could work together to behave like a car. In cognitive psychology, parallel distributed processing (or connectionism) can be thought of as a low-level modelling approach and rule-based systems can be thought of as a high-level approach. These will be considered in turn in this section.

1.1 Parallel distributed processing

Parallel distributed processing or PDP modelling (Rumelhart and McClelland, 1986), sometimes known as connectionist or neural network modelling, involves building models that match human performance by programming artificial neurons into networks. The artificial neurons with which the model is built, though far simpler than the actual neurons found in the human brain, have certain neural-like properties (such as the way activation can spread between them). Also, the structure of these networks, in terms of the number of artificial neurons and their interconnections, is far simpler than the human brain. However, a PDP model can provide many useful insights into how a neural-like architecture can exhibit human-like cognitive behaviour.

The artificial neurons that make up a PDP model are often referred to as **units** or **nodes**. These units are connected together by links to form a network. A network of four units is shown in Figure 19.1. Here, three units are providing input to a fourth unit. PDP units can be activated and can then spread activation along the specified links to subsequent units. The strength of the signal received by the next unit is determined by the **weight** of the link. In Figure 19.1, each of the links has been assigned a numerical weight (usually a value between −1 and +1). In Figure 19.1, the left-hand link has been assigned a weight of 0.6, the middle link has a weight of 0.2, and the right-hand link has a weight of 0.3.

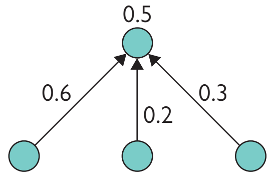


FIGURE 19.1 A network of four nodes.

Whether the input to a unit is sufficient to activate it is determined by the unit's threshold value. The threshold value is the level of input that a unit must receive in order to send an output to subsequent units in the network. For example, in Figure 19.1, the top unit receiving input from the other three has a **threshold value** of 0.5. The threshold value is shown inside the unit. (For simplicity, threshold values have not been specified for the lower three units.) As the top unit has a threshold value of 0.5, this means that an input from the left-hand link, having a weight of 0.6, would itself be sufficient to activate the unit. The middle and right-hand links *individually* would not activate the unit, as their weights (0.2 and 0.3 respectively) are lower than the threshold value. However, these two links working simultaneously can activate the unit, as 0.2 and 0.3 add up to the threshold value of 0.5.

One of the most common forms of PDP network is a feed-forward, three-layer network, consisting of an input layer, hidden layer, and output layer. An example is shown in Figure 19.4. The **input layer** of the network contains the units that receive input from the outside world. If activated, these units send outputs to nodes in the **hidden layer**, which has no direct link to the outside world. Finally, the units of the **output layer** receive input from the hidden layer and send an output to the outside world. Whether the units of each layer send a signal to the next layer is determined by the input weights they receive and the threshold value

✱ ACTIVITY 19.1

(a) Assuming that units A and B in Figure 19.2 are activated simultaneously, and both have an activation level of 1.0, what is the maximum threshold value of unit C that would still allow it to be activated?

(b) Assuming that units A and B in Figure 19.3 are activated simultaneously, and both have an activation level of 1.0, what is the minimum weight of the link from unit A that would allow unit C to be activated if it had a threshold value of 0.6?

COMMENT

(a) The maximum threshold value would be 0.8. This would be the sum of the weights of the two input links 0.3 and 0.5. If the threshold value was higher than 0.8, it would be impossible for the unit to fire given these two inputs.

(b) The minimum weight would be 0.1. This would make the sum of the two weights, 0.6, equal to the threshold value of unit C. If the weight of the link was lower than 0.1, it would be impossible for unit C to fire given a threshold value of 0.6.

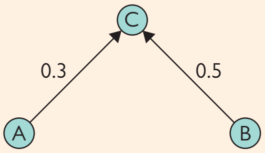


FIGURE 19.2

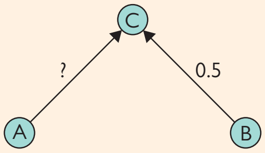


FIGURE 19.3

of the unit. The behaviour of a network (i.e. the outputs it provides according to inputs received from the outside world) will be determined by the threshold values of the units and the weights of the links.

PDP models can also 'learn' – one way in which this can be done is via an automatic procedure for successively modifying the weights of its links until the network ultimately produces the correct outputs. During **training**, the network will be provided with a large number of example inputs, and a specification of the output the network should produce for each one. For example, suppose we present a network with an input pattern and a unit in the output layer provides an output when it is not desired. This can be corrected in the model by decreasing the weights of the links that provide this unit's input, and so making it more difficult for the unit to reach its threshold value when the network is presented with the same input pattern again. Conversely, if a unit in the output layer fails to provide an output when desired, then the weight of its input links can be increased, making it easier for the unit subsequently to reach its threshold value.

In practice, a PDP model 'learns' in this way not through modifying just the weights of the links feeding directly into the output layer, but by adjusting all of the links that indirectly affect the output layer. Modifying the behaviour of a unit in the output layer can therefore involve modifying the weights of links between the hidden and output layers, and between the input and hidden layers. It is possible to adjust the weights of the network in this way automatically, by calculating backwards from the links between hidden and output layers to the links between input and hidden layers. This process of adjusting weights backward from the output layer is known as the **backward propagation of error**.

In addition to their ability to learn, PDP models have other characteristics that are important for our understanding of cognition:

1. PDP models need not contain any explicit rules, but can still behave as if they are following such rules. For example, Prince and Smolensky (1997) developed a PDP model of how words are arranged in order to produce grammatical sentences. The behaviour of the model could be explained in terms of a set of grammatical rules, but none of these rules were explicitly present in

the model. The behaviour of the model *conformed* to, but was not *governed* by, any explicit rules.

2. If a network is damaged, for example through the removal of specific units, then performance of the model will typically deteriorate slightly, rather than fail completely. This gradual deterioration in performance is referred to as **graceful degradation**. It is also a characteristic of the human brain – for example, the gradual loss of neurons in the brain due to ageing has a gradual rather than sudden effect on cognitive processes such as learning and memory.
3. PDP models can exhibit emergent properties. An **emergent property** is a behaviour that the model comes to exhibit over time, through training, that was not explicitly programmed into the model. This feature also makes PDP models interesting from a psychological viewpoint, as much of human learning derives from accumulated experience rather than explicit instruction.

1.2 Rule-based systems

PDP models can be contrasted with rule-based systems. A **rule-based** system models cognition as an explicit set of rules – for example, production rules – that provide a recipe for how the model should behave. **Production rules** contain two parts – a condition and an action. They are structured in the form 'IF condition is met THEN perform action'. The **condition** specifies what must be true in order for the rule to be applied. The **action** is what the production rule should perform if the condition is true. If a production rule matches its condition and performs its action, then the rule is said to have **fired**. A rule-based model is constructed out of a set of production rules that together can produce the desired cognitive behaviour.

(1) IF the cup just contains milk AND the tea in the pot is ready
THEN pour the tea into the cup.

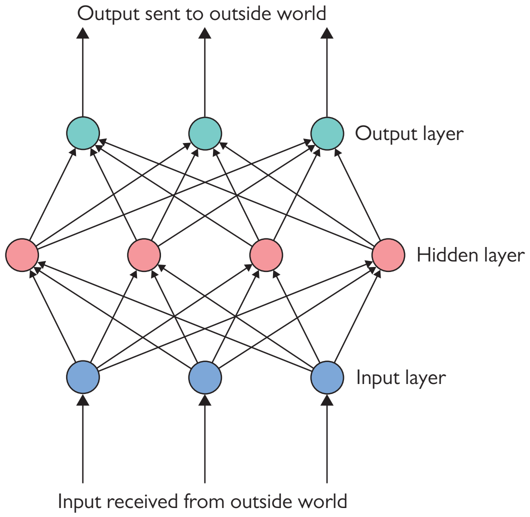

Output sent to outside world
Output layer
Hidden layer
Input layer
Input received from outside world

FIGURE 19.4 A PDP architecture comprising an input layer, hidden layer, and output layer, adapted from Rumelhart and McClelland (1986).

The production rule written in an English-like form in (1) could be used in a model of how to make a cup of tea. A complete model of tea-making might contain production rules for boiling the kettle, warming the pot, adding tea bags to the pot, adding sugar, and stirring the tea. Each production rule would fire when the current state of the tea-making process matched its condition. The rule in (1) would only fire once the milk had been added to the cup and the tea was ready to pour.

It is possible to model complex cognitive processes using production rules. For example, a rule-based model of how humans produce grammatical sentences may have production rules for selecting an appropriate verb depending on the meaning of the sentence, and for selecting an appropriate ending for the verb depending on the tense to be used.

Unlike PDP models, rule-based models focus on how the cognitive tasks humans perform can be understood as the processing of information without considering how such processing might be realized in the brain. The relationship between the approaches taken by PDP and rule-based systems can be understood in terms of Marr's (1982) levels of explanation (discussed in Chapter 1). Marr argued that psychological explanations can be understood at any of three levels: computational, algorithmic, and hardware. The traditional PDP approach places a greater emphasis on the hardware level, arguing that if the model reflects some basic properties of the human brain, then interesting psychological behaviour will emerge. On the other hand, a traditional rule-based approach places virtually no emphasis on the actual brain, effectively saying that the way the brain works is more a matter for biology than psychology. Instead, it argues that psychological phenomena can be most appropriately explained at the computational and algorithmic levels. In practice, this distinction is increasingly fuzzy; many PDP models explore complex cognitive behaviours, and current rule-based models are concerned about biological plausibility.

1.3 Cognitive architectures

Newell (1973) argued that it was not sufficient to develop a collection of discrete models to describe a broad range of psychological phenomena. Instead, there should be some integration and consistency across the models being developed. For example, playing chess, recognizing objects, and producing grammatical sentences are all psychological processes that use long-term memory. If rule-based models of these three psychological phenomena contained completely different ways of representing, organizing, and retrieving from long-term memory, this would clearly be a problem. Humans use the same cognitive processes across a range of tasks.

This led to the development of cognitive architectures that could account for a range of cognitive processes using the same modelling components. A **cognitive architecture** is an overarching framework that can account for a wide range of phenomena using a fixed set of mechanisms. As well as maintaining consistency across a set of models, cognitive architectures have another important advantage. Cognitive architectures distinguish clearly between the cognitive model and the computer (and any associated programming languages) on which the model is running. When a cognitive model is developed using a standard programming language (such as the C programming language) there is a need to distinguish which parts of the program are psychologically relevant and which are just dependent on the programming language being used. For example, a programming language has facilities for storing information but the model developer is not necessarily claiming that the way this information is stored in the computer bears any relation to the way humans store information in memory. A rule-based cognitive architecture is 'emulated' on a computer, but there is a clear distinction between the working of the model and the working of the computer. By this we mean that the cognitive architecture is used on a computer but has its own self-contained set of processes, such as production rules. It does not directly use the general purpose processes of the computer that are used to provide the computer user with all kinds of facilities from email to word processing.

PDP is itself a cognitive architecture comprising a fixed set of artificial neural mechanisms. Two of the most well known rule-based cognitive architectures are Soar (Newell, 1990) and ACT-R (Anderson and Lebiere, 1998). In the next three sections we will look in detail at ACT-R as a rule-based cognitive architecture and at some of the empirical data it has been used to model. Particular consideration will be given to the extent to which ACT-R meets Newell's (1990) goal to develop a cognitive architecture that can provide an integrated and consistent account of a wide range of psychological processes.

SUMMARY OF SECTION 1

- Cognitive modelling involves building a model of a cognitive process and then comparing the behaviour of the model against human performance.
- The two main types of cognitive modelling are parallel distributed processing (PDP) and rule-based systems.

- PDP models have neural-like properties and aim to demonstrate how a neural architecture can support human cognition.
- Rule-based models focus on how information is processed and give less consideration to how cognitive processes are realized in the brain.
- Rule-based cognitive architectures such as ACT-R and Soar attempt to provide an integrated account of a range of cognitive theories and empirical findings.

2 AN OVERVIEW OF ACT-R

To give you a taste of cognitive modelling, in this chapter we will describe, evaluate, and use Anderson's ACT-R cognitive architecture (Anderson and Lebiere, 1998). ACT-R is perhaps the most widely used cognitive architecture in the cognitive modelling community, and reflects a trend towards **hybrid models** that attempt to span Marr's levels of explanation. Although ACT-R is primarily a rule-based cognitive architecture, it has certain characteristics more usually associated with PDP.

2.1 A brief history of ACT-R

ACT (which stands for 'Adaptive Control of Thought') has its roots in Anderson and Bower's (1973) theory of human associative memory, and their model of it, called HAM. A number of different versions of ACT have been developed. The first version of ACT proper, called ACTE, combined elements of HAM's memory representations with rule-based production systems that modelled control of behaviour and more complex activities like problem solving. In 1983, Anderson revised ACTE producing ACT* (pronounced 'act star'), which revised the underlying memory system to be more plausible biologically and introduced a mechanism for learning new rules for the first time. ACT* was the first complete theory in the ACT series, and was capable of modelling a wide range of behaviours, from memory to complex problem solving and skill acquisition.

In 1993, ACT-R was developed and since then it has been gradually revised. The 'R' stands for 'rational' and refers to Anderson's (1990) theory of 'rational analysis' (recall the different explications of rationality discussed in Chapter 12). Basically, **rational analysis theory** states that each component of the cognitive system is optimized with respect to demands from the environment, given its computational limitations. ACT-R also shifted the emphasis towards a finer-grained model. Whereas earlier rule-based models tended to use a small number of complex rules, in ACT-R there is a definite shift to a larger number of simple rules. In fact, Anderson and Lebiere (1998) argue that the simple chunks of ACT-R's memory, and the simple rules of its production system, are 'atomic' in the sense that they should not be broken down into further, more fine-grained ACT-R constructs.

2.2 The architecture of ACT-R

The ACT-R cognitive architecture comprises a clear set of components, whose interactions lead to its special behaviour. These components are more or less distinct modules within it. We have shown an overview of the architecture for ACT-R in Figure 19.5 below.

The ACT-R architecture consists of three core components: **declarative memory**, **procedural memory** (see Chapter 8), and a **goal module**. Put simply, declarative

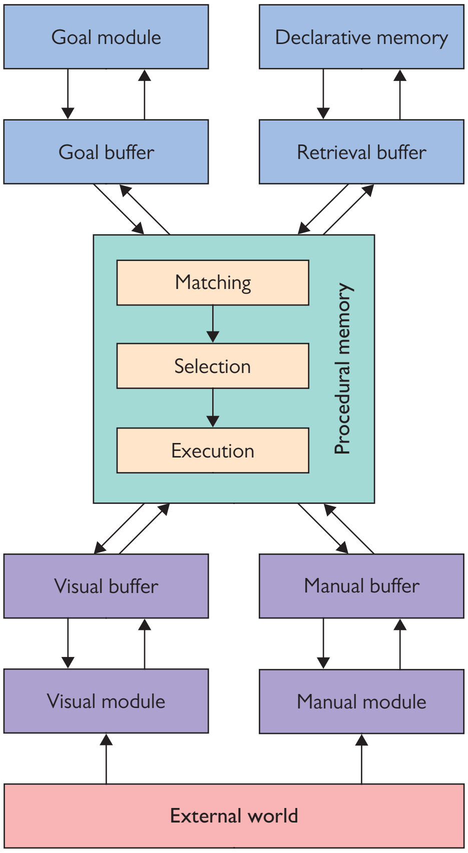


FIGURE 19.5 An overview of the ACT-R architecture (adapted from Anderson *et al.*, 2004).

memory is full of things that we just know, for example, that '1 + 3' is '4', that grass is green, and that books have pages in them. Procedural memory, on the other hand, is full of things that we know how to do. For example, few people would know the answer to '154 + 367' directly, but most would know a procedure that can be used to work out the answer. The steps in this procedure would be stored in procedural memory.

Importantly, declarative memory and procedural memory are not independent, but interrelated. There are two kinds of connection between them. First, a production rule in procedural memory often involves retrieving elements from declarative memory through the retrieval buffer shown in Figure 19.6. Second, new production rules can be created in procedural memory from 'chunks' in declarative memory (this process will be discussed in Section 2.4). This process is known as **production compilation**. The importance of production compilation can be seen in ACT-R models of learning, where new rules are formed as learners become more skilled at solving problems. For example, an intermediate chess player will pick up new rules through experience not of the rules of the game itself, which they already know as facts, but as rules about how to defend against or exploit particular situations on the board.

There is more to the ACT-R architecture than the interplay between declarative and procedural memory. There is also a **goal module** (see Section 2.5). This is important for models involving lots of rules, as the current goal is a kind of focus of current attention – this represents what ACT-R is currently trying to do.

ACT-R also has a range of (very simple) perceptual and motor components that allow it to model visual and auditory stimuli, and actions including changing visual attention and vocalizing, as well as physical actions. These components are not directly linked to procedural memory, but – like declarative memory and the goal system – are connected through small shared spaces called **buffers**. It is only the buffer contents that affect productions.

Combined, these components allow ACT-R to model a very wide range of cognitive phenomena, from memory (we will look at an example in Section 3) and problem solving (see Section 4.2) to tasks that involve perception and action (see Section 4.3).

This gives you the big picture of ACT-R. Now let's go into each part of the architecture in a little more detail.

2.3 Declarative memory

All of the ACT models, going way back to HAM in 1973, conceived of declarative memory as a collection of **chunks**, or declarative memory elements, which themselves contain a number of elements, usually between two and four. A typical chunk is shown in Figure 19.6.

This chunk encodes the addition fact, '1 + 3 is 4'. Each of the main rows in the chunk holds a value in a different **slot**. Chunks also have the 'isa' slot (pronounced 'is a'), which says what kind of chunk it is (i.e. 'this chunk isa addition fact').

```
Chunk1
        ISA addition-fact
        Number1 = 1
        Number2 = 3
        Sum = 4
```

FIGURE 19.6 An example ACT-R chunk.

Chunks are not isolated from each other; they are all linked to each other through their values in a kind of network. For example, the chunk shown in Figure 19.6 will be linked to all other chunks that are addition facts or that use the numbers one, three, or four. Chunks influence one another through these links by a limited form of spreading activation (discussed in detail in Section 3.3). Put simply, each chunk has a level of **activation**, a kind of energy attached to it. But activation in one chunk tends to leak out and add to all the other chunks it is connected to through its values. For example, increasing the activation of the addition fact '1 + 3 = 4' would automatically increase the activation of other addition facts and of chunks that contain '1', '3', or '4' in their slots. Chunks that are both addition facts and that contain '1', '3', or '4' would get even more activation.

Activation is central to many aspects of ACT-R. If a chunk has a lot of activation, it will be easy to find and to retrieve quickly from memory. If a chunk has very low activation, it will be hard to find and may never be retrieved at all. Highly active chunks, therefore, look a bit like short-term memory, and chunks with less activation look a bit like long-term memory. If a chunk has no activation, it has effectively been forgotten, at least temporarily, until its activation level is increased.

Activation is not constant for a chunk. If nothing happens, a chunk slowly loses activation. But if a chunk is used, it gains activation, and the more it is used, the more it gains. This allows ACT-R to model effects such as priming (e.g. see Chapter 6). For example, if a chunk is retrieved, its activation will increase. If the chunk is then retrieved again shortly after (before the added activation has been lost), it will be retrieved more quickly. So declarative memory in ACT-R is a lot more than a place for storing chunks. It plays an active and essential role in the behaviour of ACT-R itself.

2.4 Procedural memory

The central component of ACT-R is its procedural memory. This represents procedures in the form of production rules, which have a condition (i.e. IF) part, and an action (i.e. THEN) part. Unfortunately, since ACT-R is a computer program, its rules tend to be written in a fairly cryptic form. Box 19.1 shows a rule, first in an English-like representation, and then in ACT-R form.

To give you an idea of the relation between these representations, the condition part of the rule (the IF part) can be found in the rows before the '= =>' character, and the action part (the THEN part) appears after it. The bits that begin '= something >' are references to chunks, and each line that follows is a slot name followed by a value. When a word begins with the '=' character, it is a **variable**; that is, it can change each time the rule is used. Variables are very important to production rules, as they are what make rules sufficiently general to run with different problems. For example, the rule in Box 19.1 could be used to add any addition column for which the matching addition fact was stored in declarative memory. Variables and values will be covered in more detail later.

ACT-R uses production rules in the following way. All production rules are available to be used depending on the current state. For a production rule to be used, its condition part (the IF part) must match the current goal and use chunks already available in

BOX 19.1 An ACT-R production rule in English and in ACT-R form

Rule in English form

How to add two numbers together
IF the goal is to find the answer to *number1* + *number2* in a column
and we know that *number1* + *number2* is *sum*
THEN the answer is *sum*

Rule in ACT-R form	Commentary
Add-numbers	The production rule called Add-numbers fires if . . .
=goal >	The current goal is . . .
ISA add-column	To add up a column of an addition sum in which . . .
first-number = number1	There is a first number (=number1) to be added . . .
second-number = number2	To a second number (=number2) . . .
answer nil	And the answer is unknown (nil means it is empty).
=retrieval >	And also we can retrieve a chunk from declarative memory . . .
ISA addition-fact	And it is an addition fact stating that . . .
addend1 = number1	The first number (=number1) added to . . .
addend2 = number2	The second number (=number2) gives . . .
answer = sum	An answer (=sum)
== >	Then . . .
=goal >	In our current goal we can use . . .
answer = sum	The answer from the addition fact (=sum) as the answer for the goal.

declarative memory. This gives quite subtle control over timing. For example, the greater the activation of a chunk in declarative memory, the faster any productions that use that chunk can be fired. ACT-R uses this technique extensively to give reasonably accurate models of human response times (as we shall see in Section 3).

Activation also plays a significant role in procedural memory. Rules have activation too, and if a rule has a very low activation it might not be used even though its conditions match the current state. However, the more a rule is activated, the more likely it is to be retrieved and used again in the future.

Procedural memory is not fixed like a computer program – new rules can be learned through a process called production compilation. Production compilation is the name given to the process by which knowledge is transferred from declarative to procedural memory. The process by which production rules are learned in ACT-R has three stages:

1. Understanding. The first stage involves understanding instructions and available worked examples. For example, a teacher may provide a child with verbal instructions for multiplying two numbers together. In ACT-R this new knowledge is encoded as chunks in declarative memory.
2. Production compilation. In the second stage, we try to solve problems by applying these instructions. By working on a range of problems we start to generalize from our experience. For example, through attempting multiplication problems for themselves a child will start to realize the range of problems to which the multiplication instructions can be applied. In the initial stages, the child may only be able to correctly solve the multiplication

problem when it has certain characteristics, such as containing or not containing particular numbers. Through experience the child will come to realize how general instructions can be used to solve a wide range of multiplication problems. In ACT-R this is supported by the process of production compilation. Through application, the declarative representation of the instructions becomes transformed into a production rule. The production rule becomes more general than the declarative chunks as specific values in the chunks become replaced by variables in the production rule. For example, chunks representing the instructions for multiplying 27 by 3 may be generalized into a production rule for multiplying any two-digit number by three, and then further generalized for multiplying any two-digit number by a one-digit number. This process is described in detail in Section 4.1.

3. **Practice**. Through practice we solve problems with increasing speed and accuracy. For example, a child who has become more experienced with multiplication problems will deliberate less over the individual steps in the task, and come to solve the problems with relative ease. ACT-R explains this in terms of an increasing use of the production rules and a decreasing use of the declarative instructions. When first developed, the production rules have a low level of activation, making the retrieval and use of the production rules more difficult. Through further learning and use, the activation level of the production rules increases until they take over from the declarative representation of the instructions.

Much of ACT-R's explanatory power comes from the production compilation process, which sets up a continual interplay between the chunks in declarative memory and the rules in procedural memory. Instructions for how to perform a given task start out as chunks in declarative memory, and then as performance improves through practice, these chunks are turned (or compiled) into rules in procedural memory, which do the same thing as the chunks but do it automatically.

To complete the circle, as these rules are used, they may themselves create new chunks in declarative memory. As we shall see in Section 4.2, for example, a child who does not know from memory the answer to '4 + 2' may use a counting procedure to arrive at the answer. The child may then remember this answer and so subsequently can provide the answer to '4 + 2' directly from memory. In ACT-R this is modelled by a production rule (e.g. the child's counting procedure) producing a chunk in declarative memory (e.g. a chunk representing '4 + 2 = 6') that can later be used to answer the same question without using the original production rule. We will come back to ACT-R's approach to learning in Section 4, where we will look at a model that shows it working in practice.

The idea that memory is divided into declarative and procedural memory is central to the ACT-R theory. One of the sources of evidence for this distinction is the experiment conducted by Rabinowitz and Goldberg (1995), which relies on the fact that declarative encodings of instructions can be reversed more easily than procedural encodings of the same instructions (see Box 19.2).

2.5 Goals and the goal module

Production systems like ACT-R's procedural memory grew out of work on problem solving systems in the field of artificial intelligence. These systems adopt a goal-directed approach to problem solving. Basically, they take a current goal and a current state, and the system acts either to achieve the goal or add a new goal that needs to be completed first. For example, if someone wishes to have home-made lasagne for their meal, they will probably first set themselves the goal of assembling all the ingredients in their kitchen. This may involve going shopping. Once the initial goal of assembling the ingredients has been met, then he or she can move onto the next stage and make the lasagne using the ingredients.

In ACT-R, if the current goal (e.g. finding the ingredients) is being used, it is held in a **goal buffer** managed

BOX 19.2 RESEARCH STUDY Rabinowitz and Goldberg's (1995) experiment

Rabinowitz and Goldberg's (1995) experiment investigated differences between declarative and procedural encodings of the same instructions. If a declaratively encoded instruction can be reversed more easily than a procedurally encoded one, this lends support to the idea of relatively distinct memory systems.

Participants were given a simple alphabet arithmetic task, with stimuli like 'C + 3 = ?', where the number indicated how many letters the participant should advance along the alphabet. The expected response here would be 'F'. Two participant groups took part in the experiment. The first group was given 12 letter–number combinations in rehearsal, but the second group was given 72 (thereby gaining more practice time). After this practice time, both groups were tested on additive alphabet arithmetic, but also on a transfer task – subtractive arithmetic on the same problems, for example 'F – 3 = ?', with the expected response 'C'.

Participants who received less practice performed better on the transfer task when the problems presented featured the same number and letters as problems they had tackled in the first task. For example, they answered more quickly to 'F – 3 = ?' if they had already seen 'C + 3 = F'. This was not the case for participants who had received more practice.

Participants who received less practice could solve subtraction problems by reversing the addition solutions held in declarative memory. Participants receiving more practice had built a procedure for arithmetic addition within their procedural memory, and this could not be reversed.

by a goal module. The current goal – the goal in the goal buffer – is fundamental to the behaviour of production rules. It represents the current context of the task, and is the main source of activation for declarative memory retrieval. In ACT-R, however, the nature of goal memory is still an open question. Working memory (Baddeley, 1986) could be considered one type, but there are expected to be others (Anderson *et al.*, 2004).

One open question is the role of activation in goal selection. Norman (1981) describes a number of common types of slips and errors that can occur because of changes in activation. For example, one common error is a 'loss of activation error' where a goal loses activation, perhaps due to a delay or interruption. You might forget one ingredient for the lasagne when shopping, because so many other goals are going on (getting to the shops, taking money, paying, and so on). A second is a 'capture error' where a more frequently used goal takes over a plan. For example, if the bus you use to go shopping happens to be the same one you use every day to go to work, you may find yourself getting off the bus at the wrong stop, and setting off for work instead.

Although the current implementation of ACT-R has a goal module that only provides a goal buffer containing a single chunk, with no goal memory system, the theory is not committed to that. Trafton *et al.* (2011), for example, describe a more subtle activation-based goal system written as a replacement to the one built into standard ACT-R, which is capable of accounting for some of the slips and errors described by Norman (1981).

2.6 Perception and action

Most often, an ACT-R model is intended to be compared to a participant in an experiment, so ACT-R has a wide range of perceptual and motor modules that enable it to model the kinds of stimuli and actions used in classic psychology experiments. Typically, the model corresponds to a participant sitting in front of a computer, looking at a monitor, listening to speakers, and using a keyboard and mouse. As such, ACT-R includes standard visual and auditory location and identification modules, and motor and speech modules (allowing sub-vocalization, so that rehearsal can be modelled), but it is designed so these can be replaced by more complex versions if needed. Each module is connected to the central production system through its own buffers, allowing rules to pick up perceptual information and start appropriate actions.

Perception and action are relative newcomers to ACT-R. Earlier versions focused primarily on the relationship between procedural memory, declarative memory, and goals. ACT-R has evolved to incorporate perception and action more deeply into its models, yet it does not claim to model perception or action processes themselves.

Visual perception, for example, is immensely complex, as optical patterns on the retina are interpreted and transformed into cognitive representations. ACT-R, as it stands, does not attempt to explain any of those processes, although it does describe some of the common perceptual effects. This distinction between explaining cognitive phenomena and simply describing them is an important one: both may be useful, but to a cognitive psychologist, explanation is far more useful and interesting.

ACT-R's visual location module, for example, models pre-attentive visual processing effects such as 'pop out', where a target can be found rapidly among an array of distractors when the target differs from the distractors by a unique visual feature, e.g. finding one green object in a field of blue ones. The time to complete such a search tends not to increase even as the number of distractors increases. However, when the target is uniquely specified by a *combination* of features, e.g. finding a green square in an array of distractors, some of which are green circles while others are blue squares, then search time tends to increase linearly in relation to the number of distractors. This module therefore allows an ACT-R model to make more accurate predictions for tasks. ACT-R does not explain how pre-attentive visual processing happens, but it does show how it can play its part in a more complex task.

Consider the Towers of Hanoi problem you encountered in Chapter 10. Many effective strategies for the Towers of Hanoi use perception to avoid having to remember too many sub-goals, by making the current state of the problem easily accessible to production rules. Without modelling perception, these strategies simply cannot be modelled.

While accounting for perception and action is important to an accurate model of the behaviour of a participant in front of a computer, what is happening inside these standard modules is in most cases trivial, and not especially important to the rest of the model. Accordingly, there are clear areas where ACT-R does not claim any psychological plausibility.

SUMMARY OF SECTION 2

- ACT-R is a cognitive architecture based around three core components: declarative memory, procedural memory, and a goal buffer.
- Declarative memory is organized as a set of interconnected chunks forming a network, each chunk having a level of activation.
- Procedural memory is comprised of production rules that also have activation levels and that can perform some action if the condition part of the rule is met.
- ACT-R production memory is not fixed; new rules can be learned, modelling skill acquisition through instruction, examples, and practice.
- Coordination between production rules is managed by a current goal held in a goal buffer, representing the context of the task.
- In its standard implementation ACT-R rules depend on a single current goal, and do not model the effects of activation on goal selection.
- Because most real cognitive behaviours depend on an external world, ACT-R includes modules that enable models to incorporate perception and action.

3 ACT-R ACCOUNTS OF MEMORY PHENOMENA

As a demonstration of the ability of ACT-R to model human memory, Anderson *et al.* (1998) developed a single model of human performance on list memory tasks. List memory is an experimental paradigm used in cognitive psychology to investigate how people store and recall items from short-term memory. Typically, participants in an experiment are presented visually with a list of items (such as words or numbers) one after another. They are then asked to recall the presented items, possibly after some delay. A restriction may be placed on the order in which the items should be recalled. The participants may be requested to recall the items in the precise order in which they were presented (termed forward recall), in reverse order (termed backward recall), or in any order (free recall).

The ACT-R model of list memory nicely illustrates many key features of the ACT-R cognitive architecture, and list memory has been an active area of ACT-R research in recent years. Although the list memory task is highly artificial, the precise nature of the task and the wealth of empirical data and theoretical explanations of results (e.g. Baddeley, 1986) provide a lot of information to support the building and evaluation of ACT-R models.

Here we will focus on how ACT-R models human performance of forward recall. The human criteria against which the model will be compared are recall latency (i.e. time taken to recall an item) and recall accuracy. Accuracy and latency data in recall for a nine-element list are shown in Figure 19.7. In the empirical study (Anderson *et al.*, 1998) from which these data were collected, participants were initially presented with a string of empty boxes, one to contain each item in the list. Participants were therefore aware of the list length from the beginning of the study. The items were then presented in their appropriate box, one at a time. As one item appeared, the previous item disappeared, so that only one item was visible at any one time. As the last item disappeared, subjects were instructed either to recall the items in a forward or backward direction. Participants would then enter the items into the boxes in order, and could not return to an earlier box once it had been visited.

If we first look at the graph for recall accuracy, two important features should be noted. First, recall accuracy is highest for the first elements in the list. This is called the primacy effect and can be explained by assuming that participants rehearse the first elements of the list during the presentation of the later items. Second, accuracy is higher for the last element than the two preceding elements. This is called the recency effect and is thought to be due to the last item still being accessible from memory during the recall phase,

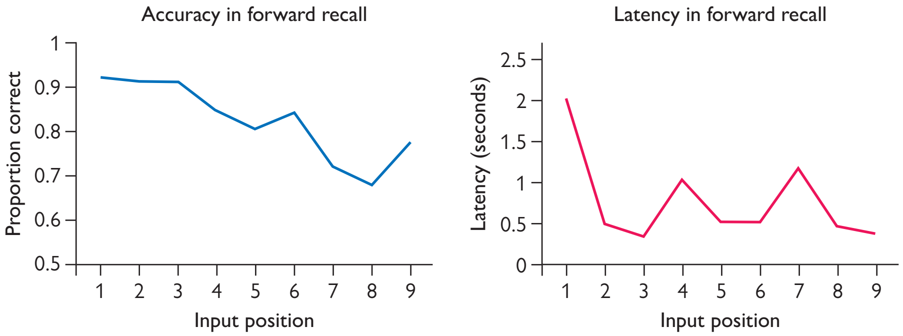


FIGURE 19.7 Recall accuracy and latency in the forward recall of a nine-element list.

even though it has not been rehearsed, as its activation level has not decayed. Turning to the graph of recall latency, it should be noted that recall is slower for elements one, four, and seven. This is conjectured to be due to the way the items are chunked in declarative memory.

In order to accurately reflect the empirical data, the ACT-R model needs to:

(a) have a representation of how items are chunked in declarative memory

(b) have production rules for the rehearsal of items and retrieval from memory

(c) model activation levels and show how they affect recall accuracy and latency.

These three points will be considered in the following three subsections.

3.1 Declarative representation of lists

Within the ACT-R model of list memory, the list itself is represented in declarative memory as chunks. Chunks are used to represent a list as a set of groups. A **group** can contain as few as two and as many as five or six items.

The way that people mentally group telephone numbers could be represented as ACT-R chunks. Consider for example the main switchboard number for The Open University in Milton Keynes. Written without any spaces to indicate groups the number is 01908274066. Some people, particularly those familiar with the Milton Keynes dialling code, will group the first five items (01908). Individual differences are found in how people tend to group a six-digit telephone number, either as two sets of three, or as three sets of two. This gives two common groupings of the number, either into three (01908 274 066) or into four (01908 27 40 66) groups.

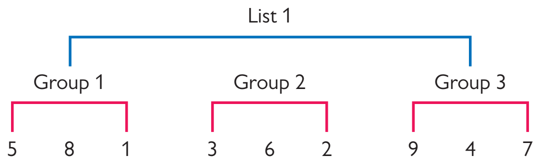


FIGURE 19.8 Organization of a nine-element list into three groups of three.

In a list recall task, participants often organize the list into three groups. This is found to optimize the number of items that can be remembered. For example the list 581362947 is often grouped as shown in Figure 19.8. To distinguish it from any other lists, we shall refer to it as List1. The three groups are referred to as Group1, 2, and 3. The model of Hitch *et al.* (1996) also represents a list as a set of groups, each containing a small number of items.

When grouping lists in declarative memory, one chunk is associated with each group and each individual item within the group is associated with a chunk. List1, as grouped above, would be encoded using 12 chunks, one chunk for each of the three groups and one chunk for each of the nine elements. A chunk can therefore hold a group of items or just a

✱ ACTIVITY 19.2

How many chunks in total are required to represent a list of 15 elements with:

(a) a group size of three

(b) a group size of five?

COMMENT

(a) Twenty chunks would be required. Five chunks are required for the five groups of three elements. Fifteen chunks are required for the individual elements. (Note that the list itself does not have a chunk. 'List1' appears as a value in each of the 15 chunks.)

(b) Eighteen chunks would be required. Three chunks are required for the three groups of five elements. Fifteen chunks are required for the individual elements.

Chunk1

Position = 1st

Group = Group1

List = List1

Content = 5

FIGURE 19.9 The chunk associated with the first item of List1.

single item. (Recall from Chapter 10 that Chase and Simon (1973) found that expert chess players could remember the positions of more chess pieces after a glance than novices. This suggests that, although experts and novices can form the same number of chunks in short-term memory, experts can represent the positions of a greater number of pieces in one chunk.)

As mentioned in the previous section each chunk is represented using **slots** and **values**; however, the chunks used to encode groups and individual items have a slightly different set of slots. A chunk associated with an individual element has slots for the group to which it belongs, its position within the group, the overall list to which it belongs, and its content (i.e. the list item). The chunk associated with the first item in the list is shown in Figure 19.9. This chunk states that the item in the first position of Group1 of List1 is the value '5'. It is important to note, according to the assumptions of ACT-R, that each individual element has a slot associating the element directly to the list as well as to the group to which it belongs. This allows ACT-R to model some complex memory effects, as we shall see later in Section 3.3.

A chunk encoding a group has slots to indicate the list to which it belongs, the number of elements in the group, and the position of the group within the list. Each chunk also has a unique name by which it can be referenced. Figure 19.10 represents the chunk associated with the first group of List1. (This group has been given the name Chunk10. The chunks associated with the nine elements of List1 have been named as Chunks 1 to 9, and the chunks associated with the three groups have been named as Chunks 10 to 12.)

3.2 Production rules for the rehearsal and retrieval of lists

The ACT-R model has to perform a number of procedures in order to simulate performance in the list memory task. The list contents have to be rehearsed in memory and then retrieved, and responses have to be given. These procedures are stored in the form of **production rules** in procedural memory, as described earlier.

The condition part of the production rule may specify what must be the current goal, and may specify what chunks have to be made available in declarative memory. For example, as was shown in Box 19.1, the Add-numbers production rule could only fire if

Chunk10

list = List1

Size = 3

Position = Group1

FIGURE 19.10 The chunk associated with the first group of List1.

✱ ACTIVITY 19.3

(a) Write down the slots and values of the chunk associated with Group3 of List1.

(b) Write down the slots and values of the chunk associated with the last item of Group2 of List1.

COMMENT

(a) According to our numbering scheme, the chunk associated with the third group is Chunk12 and it has a size of three. It also maintains a link to the list. The chunk would therefore look like Figure 19.11.

(b) The last element of Group2 would be Chunk6. As this chunk refers to an individual item, it has a slot for its content. The sixth number in the list is 2. The chunk would therefore look like Figure 19.12.

Chunks therefore organize a seemingly flat list of elements into a **hierarchy**, where the list is initially broken down into three groups and the groups are then broken down into individual elements.

Chunk12

List = List1

Size = 3

Position = Group3

FIGURE 19.11

Chunk6

Position = 3rd

Group = Group2

List = List1

Content = 2

FIGURE 19.12

(2) Rule in English form

Get next item from group

IF the goal is to get element at position *X* from group *Y* and we can retrieve the element *Z* at position *X* from group *Y*

THEN create a sub-goal to output the item and modify the current goal to look for item *X* + 1.

the add-column goal was the current goal and the appropriate addition-fact chunk was in declarative memory. The action part of the production rule either transforms goals held in memory or performs an action to the world outside the model, such as typing a retrieved word onto a computer screen. The ACT-R model of list memory comprises a number of production rules for rehearsing, retrieving, and giving responses to the questions or tasks of the experiment. For example, (2) represents the English form of the production rule for getting the next item from a group (the ACT-R textual syntax has been removed for clarity). Basically, the rule states that if you can retrieve the item you are trying to recall (i.e. the item at position *X* in group *Y*), then set a sub-goal to output (i.e. say) the retrieved item and then move onto the next item (i.e. item $X + 1$).

(3) shows the production rule used to output an item in a recall task via a key press. Here the action part just provides an output and does not transform any goals.

(3) Rule in English form

Type out the item

IF the goal is to output item *X* and we can retrieve a key for *X* that is found on the keyboard

THEN output by pressing the key.

The production rules shown above contain certain characters written in italics (*X*, *Y*, and *Z*). These are variables that act as empty slots and can accept a range of specific values. These are used in order to satisfy the ACT-R assumption that production rules should provide a level of generalization and be applicable to a range of specific cases. For example, the 'Get next item from group' production rule can be used to get any item from a group. Similarly the 'Type out the item' production rule can be used to output any printable character. Goals will be discussed in more depth in Section 4 on arithmetic skills.

3.3 List activation

As we discussed in the previous section, retrieval from declarative memory is performed within the condition part of the production rule. The success of this retrieval process for recall tasks is a function of the activation level of the chunk that matches the condition. The higher the activation level of the chunk, the more easily – and faster – it can be retrieved.

Activation is part of the PDP-like nature of the ACT-R architecture. Activation of (artificial neuronal) elements is used within PDP architectures and is itself inspired by our understanding of neurology. The activation level of each chunk in declarative memory is calculated using a set of activation equations provided and modifiable within ACT-R.

The activation level of a chunk is calculated as the sum of its base-level activation and associative activation. The base-level activation of a chunk depends on the number of times it has been rehearsed in memory and the amount of time that has elapsed since it was last rehearsed. If a chunk has been rehearsed a high number of times and only a small amount of time has elapsed since the last rehearsal, then its base-level activation will be high.

The base-level activation provided by the ACT-R architecture can help to account for primacy and recency effects in list memory. The primacy effect is due to the number of times the item has been rehearsed, increasing its base-level activation. The recency effect is due to the small time lapse since the presentation or last rehearsal of the item, which also means base-level activation will be high.

The association strength is the strength of the bond between an item and the required chunk, and influences the flow of activation between chunks. Looking back to Figure 19.8, there will be an association strength between item '5' on the list (the first number on the list) and a chunk that encodes that item, such as Chunk1. The strength of association between an item and a chunk depends on the total number of associations that the item has. An important assumption of the ACT-R architecture is that activation is a limited resource. If an item is only associated with one chunk, then this chunk receives the full associative strength of the item and, therefore, the full effect of any activation. If the item is associated with three chunks, then the association strength is split three ways and less activation will flow to each individual chunk.

The limited capacity of association strength can be used to explain what is known as the fan effect (Anderson, 1974). The fan effect is the empirically observed finding that the greater the number of facts related to some concept that a subject has to memorize, the slower the subject will be to recall any one of them. For example, imagine you are asked to memorize three facts about a pretend person called Fred – that he is six feet tall, has a beard, and works in a hospital. Your recall will be slower for the fact that Fred is six feet tall because you have been asked to remember other facts about him. If you had been asked to remember just this one fact about Fred, your recall would be quicker. Although the fan effect can be explained in terms of spreading activation, the mechanism within ACT-R is precise and restricted – activation only spreads to the immediate neighbours in the network. However, spreading activation is assumed to encompass a wider region of neurons (or units) than just the immediate neighbours of the activation source.

The same interpretation for the fan effect also applies to list memory. In Section 3.1 it was shown that each chunk encoding either a group or an individual item in a list has an association to the list itself, in that case, List1. The association strength for List1 has to be shared out among all the associated chunks. The more chunks that are associated with List1 (either due to a smaller group size or larger list size), the more thinly the association activation has to be spread between the chunks, making it increasingly difficult to (quickly) retrieve any of the associated chunks. Limited association strength therefore offers an account of how list size affects the recall of items from a list. The larger the list, the smaller the percentage of items successfully recalled. This is also one of the reasons why the group size is optimally set to three rather than two, as this reduces the number of associations with the list concept itself, without overloading any particular group.

In order for a chunk to be retrieved at all, its activation (which is the sum of the base-level activation and the association activation) has to reach a certain level, specified in the ACT-R model. This pre-set level is called the **activation threshold**. A chunk that falls below the activation threshold is unavailable for retrieval by the production rules. The activation level therefore affects recall success. An equation in ACT-R also specifies the relationship between activation and latency. The weaker the activation, the slower the recall process will be.

However, retrieval is not always 'all or nothing'. It is possible to select a chunk that only partially matches the item sought in the condition part of the production rule. This process, called **partial matching**, happens if, for example, the partially matching chunk has a high level of activation and a fully matching chunk is either absent or below the activation threshold. Partial matching is also important in modelling certain empirically observed effects. **Positional confusions**, where the participant recalls a correct item but in the wrong position, are common in list recall data. (Positional confusions are also discussed in Section 4.1 of Chapter 9, where they are referred to as transposition errors.) For example, in the case of List1, a participant might recall the number 4 in the seventh position rather than the eighth. Once again, equations in ACT-R determine the likelihood of positional confusion. Items are more likely to be confused if they appear in the same group, and if they appear in adjacent positions in the same group. ACT-R uses this mechanism to successfully model the positional confusions exhibited by human participants.

3.4 Running the model

We have just covered three important features of ACT-R. First, we have seen how lists can be represented in declarative memory. Second, we have seen how production rules can be used to retrieve items from memory. Third, we have seen how ACT-R employs a model of activation that influences recall accuracy and latency. Now we will consider the running of the ACT-R model of list memory to see how these features work. Figure 19.13 shows the performance of the ACT-R model on forward recall superimposed on the empirical data presented in Figure 19.7.

The results for recall accuracy from the empirical study and the running of the ACT-R model are very similar (Figure 19.13, left). Both show a primacy effect, having the highest recall for the first three elements. Both also show a recency effect for the last item

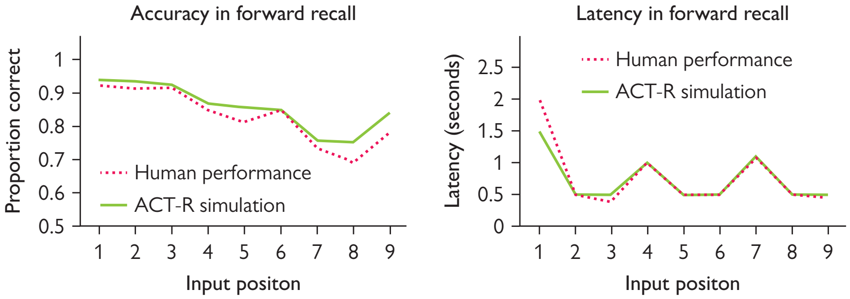


FIGURE 19.13 A comparison of ACT-R simulation and human performance on accuracy and latency in forward recall.

in the nine-element list. The simulation closely mirrors the findings but certain parameters had to be set within the ACT-R model in order to fit the results so closely. This included the parameter that affects the likelihood of a positional confusion. The issue of setting parameters in ACT-R and its implications for the validity of the model is considered in Section 3.5.

When examining latency (Figure 19.13, right), data from the empirical study show spikes at intervals of three items that are mirrored in the ACT-R model. In ACT-R, when a production rule is retrieving the next item from a list it takes longer if that item is from the next group rather than from the same group. If the next item is from the same group then the production rule just needs to increment the counter in the current group and retrieve the item. If the next item is from the next group then production rules have to be used to retrieve the next group and then retrieve the first item of that group. So, these data support the assumption that items in declarative memory are grouped, and that groupings of three tend to be used for list recall experiments.

3.5 Evaluation of the ACT-R approach to modelling memory

ACT-R has been used to develop versions of a model that can explain a range of empirical data related to list memory. The versions, however, are not identical in every way. The overall model described above has been used to model data from forward recall, backward recall, and also free recall, but in each case the model needs to be customized in certain ways to fit the data.

Certain parameters need to be set, particularly those relating to the activation settings of ACT-R within the model. For example, parameters affecting the likelihood that an item will be recalled and the time delay involved in accessing a new chunk vary across the different versions of the model. Anderson *et al.* (1998) admit this variation and offer explanations, but it is clearly an area of concern requiring further work. If such variation is required to explain different list memory experiments, which use a heavily restricted and artificial task, to what extent can ACT-R hope to provide a unified theory of cognition?

However, ACT-R does impose certain architectural constraints, limiting how the model can work. These include the procedural–declarative distinction and the use of chunks to group items in declarative memory. However, these constraints still allow for some flexibility when modelling empirical data. Decisions on how to deal with this flexibility are called auxiliary assumptions. An **auxiliary assumption** is made on a case by case basis to deal with the peculiarities of a particular experiment. How the rehearsal of previously presented items in a list occurs and how this competes with attention to the presentation of the remaining items in the list is an example of an auxiliary assumption. These can be contrasted with architectural assumptions. An **architectural assumption** (such as the procedural–declarative distinction) makes a general claim as to the nature of human cognition, and is consistent across all cognitive models developed using the architecture. If an architectural assumption of ACT-R made it impossible to model certain empirical data, then this would suggest that the assumption does not reflect a general feature of human cognition, and that the architectural assumption should be rejected or at least modified.

There is ongoing debate as to whether ACT-R (and other cognitive architectures) sufficiently constrain the modelling process. However, one advantage of ACT-R and other cognitive architectures is that at least all assumptions are made explicit, allowing such debates to occur.

SUMMARY OF SECTION 3

- Chunks are used to organize items in declarative memory into groups.
- Chunks have slots and values and are used to encode both groups and individual items.
- Chunks have an activation level comprising base-level activation and association strength.
- ACT-R is used to model the accuracy and latency of forward and backward recall, as well as other list memory experiments.
- Model fitting is used to match the ACT-R model to the empirical results.
- Any model makes associated architectural and auxiliary assumptions that jointly specify the model and how it works.

4 LEARNING AND PROBLEM SOLVING

In the previous section, we saw ACT-R's model of list memory, which focused particularly on how declarative memory items are represented and how their retrieval is affected by levels of activation. In this section, we will focus more on production rules in ACT-R and their role in the modelling of problem-solving behaviour and the learning process as a novice acquires expertise through practice.

4.1 Production compilation

As we discussed in Section 2.4, the ACT-R approach to learning comprises three stages. In the first stage of learning, instructions that the learner has been given are encoded as chunks in declarative memory. A separate chunk is used to represent each step in the instructions. Chunks that represent steps in a process (rather than facts) are called **dependency goal chunks**. A dependency goal chunk is created every time a goal has been successfully completed. For example, if the goal is to find the answer to '3 + 4', and this is solved by matching the goal with the addition fact '3 + 4 = 7', then a dependency chunk would be created. This dependency chunk would in effect say:

> *If the goal is to find the answer to 3 + 4 and there is an addition fact 3 + 4 = 7 then the answer is 7.*

This dependency goal chunk is shown to the left of Figure 19.14. The other chunks that it refers to in declarative memory are shown to the right of the figure. Any dependency chunk represents how an unsolved goal can be turned into a solved goal by using one or more other chunks in memory. The

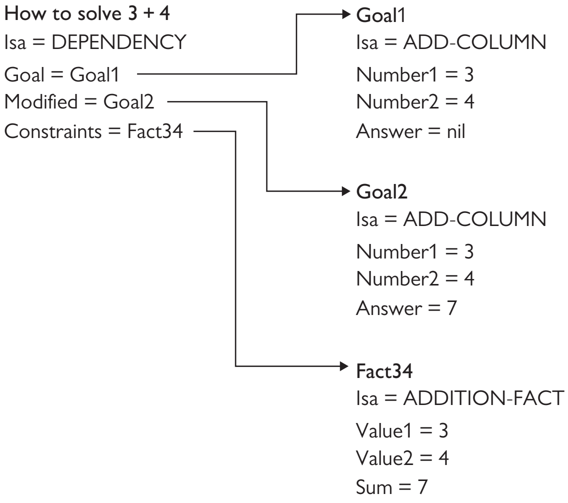


FIGURE 19.14 An example dependency for adding numbers in arithmetic.

dependency goal chunk in Figure 19.14, which for clarity we have labelled 'How to solve 3 + 4', states that the unsolved 'Goal1' (3 + 4 = ?) was turned into the solved 'Goal2' (3 + 4 = 7) using 'Fact34' (3 + 4 = 7).

Each dependency goal chunk can be thought of as representing a lesson learned from experience, and the lessons represented in dependency goal chunks are very specific. The dependency goal chunk in Figure 19.14 describes how to solve '3 + 4' and no other addition sum. Dependency goals become generalized through production compilation. The first step of production compilation turns a specific dependency goal chunk into a specific production rule. The second step turns the specific rule into a general rule. These two steps will be considered in turn.

First, production compilation works by turning the dependency goal chunk into a new production rule. This produces a new rule that has the unsolved goal and the chunks that were used to solve the goal as the IF part of the rule. The solved goal is placed into the THEN part of the rule. If applied to the goal dependency chunk shown in Figure 19.14, then the rule in Figure 19.15 would be produced. For illustration, the production rule has been given the name Add-3-to-4.

So far so good, but this produces a rule that would only work for the problem '3 + 4'. The second part of product compilation involves generalizing the rule so that it can be used to solve a wider range of goals. ACT-R achieves generalization by replacing specific

Rule in ACT-R form	Commentary
Add-3-to-4	The production rule called Add-3-to-4 fires if…
=goal>	The current goal is…
ISA add-column	To add up a column of an addition sum…
first-number 3	And the first number is 3…
second-number 4	And the second number is 4…
answer nil	And the answer is unknown (nil means it is empty).
=retrieval>	And also we can retrieve a chunk from declarative memory…
ISA addition-fact	And it is an addition fact stating that…
addend1 3	The number 3 added to…
addend2 4	The number 4 gives…
answer 7	An answer 7.
==>	Then
=goal>	In our current goal we can put…
answer 7	The answer as 7.

FIGURE 19.15 A specific production rule produced from the dependency goal chunk shown in Figure 19.14.

items in the rule by variables. The heuristic used by ACT-R is that if the same item appears in two or more places within the rule then these should be replaced by a variable. For example, in Figure 19.15, the item '3' appears in two places. It appears as the first number in the goal, and as addend1 in the addition fact. Production compilation would replace these with a variable. This means that the items in first-number and addend1 would have to be the same, but not necessarily '3'. A similar transformation would happen to the two occurrences of '4' and the two occurrences of '7'. This would in effect produce a rule that said:

> *If the goal is to find the answer to X + Y and there is an addition fact X + Y = Z then the answer is Z.*

The rule that emerges from this process would work just as well for '2 + 2 is 4' and '5 + 2 is 7' as it does for '3 + 4 is 7'. The newly made rule, created by generalizing the rule shown in Figure 19.15, is the one we saw earlier in Box 19.1. So, ACT-R does not have to be programmed to deal with every eventuality; it can learn new rules to deal with new examples.

Of course, generalizing a rule like this may not always work properly. Although the new rule has become part of procedural memory, its use depends on its activation. If the rule works, it will gain activation and will become an active player in problem-solving actions. On the other hand, if it doesn't work it will gradually lose activation and become less significant, until it is eventually forgotten.

Overall, therefore, production compilation transforms a specific 'lesson learned' represented as a dependency goal chunk into a general 'lesson learned' represented as a production rule in procedural memory.

4.2 An example of human problem-solving behaviour: addition by counting

This gives you a pretty good idea of how a single step in problem solving can be learned, but perhaps this is focusing on too small a part of the problem. Now let's look at how this can be extended to learn a more realistic skill, addition by counting. In looking at this, we are turning the clock back a bit from the examples we have shown before, to the point where we have a child who does not yet know that '3 + 4 is 7'. In ACT-R terms, there is no addition-fact for '3 + 4 = 7' in declarative memory.

This introduces a whole new issue: how do children learn these addition facts? One possibility is that they are simply learned by rote – children recite them often enough that they become chunks in declarative memory directly. However, there is a second possibility: that they are developed through a procedure. Basically, given the problem: 'What do you get if you add 3 and 4?', a child starts from '3' and adds '1' to it four times. The result is, of course, '7'. Siegler (1988) provides an account of how children use this counting strategy when learning arithmetic.

The full ACT-R model for this would be a bit hard to read and to understand, so let's look at a formatted version of it (Box 19.3). Above each rule is a plain English description of what it does.

These rules may look clumsy but they are enough to model children working out the answer to problems like '3 + 4 is 7', as if by counting on their fingers. Let's look at how these rules work in a little more detail.

First of all, the rules 'Add-numbers' and 'Sub-goal-counting' work together. Basically, if the child already knows the answer (for example, the answer is a chunk in declarative memory, as an addition fact, for example) then the child can just say it (that's what 'Add-numbers' does). On the other hand, if the child does not know the answer, they have to begin to work it out. The three rules 'Start-count', 'Add1-count', and 'Stop-count' all work together, once counting is started by the rule 'Sub-goal-counting'. 'Start-count' starts the counting process, then 'Add1-count' is used once for each step, and 'Stop-count' stops it when it is complete. Then the rule 'Found-sum' retrieves the answer from the counting process. Finally, there is a rule 'Say' that is used when the child says the answer to the problem. Table 19.1 shows how a model might trace through these various rules in practice, given the question 'What is the answer to 3 + 4?'

BOX 19.3 Production rules for addition by counting (based on Anderson and Lebiere, 1998)

This is the rule from Box 19.1 again. It can answer any addition problem for which it can access the appropriate fact from declarative memory.

Add-numbers

IF goal is to answer a question about the sum of *N1* and *N2* and can retrieve as a fact the sum of *N1* and *N2*
THEN replace current goal with one to say the sum

If Retrieve-sum cannot fire (the sum cannot be retrieved) because the addition fact is not in declarative memory, then this rule will create a sub-goal to calculate the answer by counting.

Sub-goal-counting

IF goal is to answer a question about the sum of *N1* and *N2*
THEN create sub-goal to calculate the sum of *N1* and *N2*
and create sub-goal to say the sum

If the goal is to calculate the answer by counting, then this rule turns the goal 'Add *X* to *Y*' into the goal 'Add 1 to *X*, *Y* times'.

Start-count

IF goal is to calculate the sum of *N1* and *N2*
THEN create sub-goal to add 1 to *N1*, *N2* times

This can add 1 to the number and create a new goal. For example, 'Add 1 to 4, 3 times' can lead to the goal 'Add 1 to 5, 2 times'.

Add1-count

IF goal is to add 1 to *N1*, *N2* times
and *New N1* is 1 more than *N1*
and *New N2* is 1 less than *N2*
THEN change goal to add 1 to *New N1, New N2* times

This stops the counting process when no more '1's need to be added. For example, 'Add 1 to 7, 0 times'.

Stop-count

IF goal is to add 1 to a number, zero times
THEN mark goal as completed

This retrieves the answer from the counting process.

Found-sum

IF goal is to find the sum of *N1* and *N2* and can retrieve the calculated sum of *N1* and *N2*
THEN mark goal as completed

This prints the answer on the computer screen.

Say

IF goal is to say *N*
THEN output *N* to the screen
and mark current goal as completed

However, if the same question is put again immediately, you might get a completely different behaviour from ACT-R (as shown in Table 19.2).

In practice, this general procedure can work out lots of addition facts, given the ability to count up from one. Over time, this means that chunks corresponding to these addition facts will be stored in declarative memory, and the more they are used (in other, more complex arithmetic problems, for example) the more they gain activation as chunks in their own right, and the rules may be used less and less. The example shown here, where the answer is retrieved directly after only one trial, is a bit unrealistic – in practice one might want ACT-R to go through the process many times before the new chunk can be retrieved reliably.

Of course, this is not necessarily the only way that children learn these arithmetic facts. Although some children do seem to use this process (Siegler, 1988), others may learn them by rote. Modelling can help to reveal the implications of these strategies in a way that makes it easier to design experiments and use other techniques to study them in more detail. This leads to

TABLE 19.1 ACT-R 'trace' for 'What is the answer to 3 + 4?' (first time around)

ACT-R trace	Explanation
Cycle 0: Sub-goal-counting	Create a sub-goal to find the answer to three plus four by counting
Cycle 1: Start-count	Set the goal to add one to three, four times
Cycle 2: Add1-count	Add one to three to make four and change the goal to add one to four, three times
Cycle 3: Add1-count	Add one to four to make five and change the goal to add one to five, two times
Cycle 4: Add1-count	Add one to five to make six and change the goal to add one to six, one time
Cycle 5: Add1-count	Add one to six to make seven and change the goal to add one to seven, zero times
Cycle 6: Stop-count	Goal is add one, zero times, therefore the counting goal is completed
Cycle 7: Found-sum	Calculated answer has been retrieved
Cycle 8: Say	Output the answer to the screen
'seven'	The outputted answer is seven

one of the more interesting features of cognitive models – they correspond more to individuals than to populations, and for this reason we sometimes need to be careful when comparing a model's predictions with statistical results from an experiment. In the case of the addition by counting example we have just looked at, we can study differences between individuals in problem solving by looking at the differences between the rules that represent their problem-solving strategies.

TABLE 19.2 ACT-R 'trace' for 'What is the answer to 3 + 4?' (second time around)

ACT-R trace	Explanation
Cycle 9: Retrieve-sum	Retrieve the answer to three plus four
Cycle 10: Say	Output the answer to the screen
'seven'	The outputted answer is seven

4.3 Learning in an everyday context: the development of mobile phone texting skills

The need to model a more gradual retrieval of new chunks, and therefore a more gradual increase in speed of performance, can be seen in an ACT-R model of the development of texting proficiency. Das and Stuerzlinger (2007, 2008) developed ACT-R models to simulate how proficiency in texting with a mobile phone gradually increases over many hours of practice. Understanding the longer-term use of different mobile phone keypads is of interest to researchers and also product designers. However, investigating this empirically is extremely difficult for a number of reasons. First, particularly in some parts of the world, finding novices with no prior experience of texting is becoming increasingly difficult. Second, longitudinal studies of the development of texting and other skills are logistically hard, requiring the same participants to attend a number of sessions over an extended period of time. For example, an empirical study of different computer keyboards required 20 sessions each of 45 minutes (MacKenzie and Zhang, 1999). Third, those participating in longitudinal studies of texting report boredom or fatigue (Butts and Cockburn, 2002), which can be expected to affect the results. Fourth, it may be difficult to determine how the results of an experiment translate to a slightly modified keypad design and repeating the experiment for each new phone would be close to impossible.

This had led researchers to investigate how ACT-R could be used to simulate how users become more proficient at texting. This would allow predictions to be made for new keypad designs. Feedback from the simulation could inform rapid improvements to the design without the costs or time delays demanded by a longitudinal study. A similar approach could be potentially used to simulate the long-term use of other devices such as a car dashboard.

The model of texting that Das and Stuerzlinger (2007, 2008) developed made use of the visual and motor components of ACT-R. The model therefore encompassed not only cognitive processes involved in

✱ ACTIVITY 19.4

Why do you think ACT-R uses a different process to give the answer second time around?

COMMENT

The second time ACT-R runs, the chunk that records the answer to '3 + 4 is 7' has already been stored in declarative memory, and has a relatively high activation. This means that second time around, the rule 'Add-numbers' can pick up the answer directly.

texting but also perception of the environment and interaction with the keyboard.

To inform the model of interacting with the keyboard some informal sessions were run with novice users in order to determine how a mobile phone is typically operated during texting. This led them to devise a layout similar to that shown in Figure 19.16. When a right-handed person was operating a phone, they would tend to start with their thumb on the start position, move their thumb to the relevant key, press it the required number of times, then move their thumb to the recoil position before moving onto the next key. Inexperienced users in particular tended to move their thumb to the recoil position between key locations in order not to obstruct their view of the keyboard when searching for letters.

Production rules representing different thumb movements were then specified using the ACT-R motor module. These represented individual movements such as moving to a key and pressing it once, and pressing again on a key that the thumb is already positioned over. The ACT-R motor module could estimate timings for each of these actions. Similarly, the vision module represents the user's view of the phone and incorporates procedures representing, for example, how to locate a particular letter on the keypad.

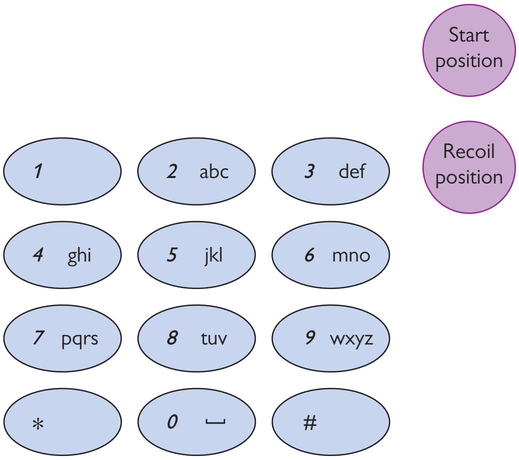


FIGURE 19.16 Typical thumb positions when texting on a mobile phone.

Based on previous empirical evidence, learning and therefore speedier performance is modelled as a gradual shift over time from locating where letters are on the keypad during texting to remembering their position. The expert's reduced need to visually scan the keyboard for letters accounts for their increased speed. A similar effect can be found in the use of a QWERTY keyboard by touch typists.

This trend in performance is modelled in ACT-R in a way that is similar to the addition example in the previous section. The user has the goal of locating a particular letter on the keypad. If the location is accessed from memory, then the user can proceed to press the key. If not, then the user has to scan the keypad to locate the correct key. This is essentially the same ACT-R process by which a child will solve an addition sum by counting if the answer cannot be accessed from memory.

In order to model a more gradual retrieval of new chunks than found in the addition example, an equation was added to the ACT-R model to slowly vary the activation level of the new chunk. This varies the activation level (and therefore probability of successful recall) depending on the amount of practice and number of previous recalls of the letter position. This prevents an instantaneous jump from not knowing to knowing the location of a letter. The trend in performance found with the ACT-R model closely mirrored previous empirical data. The ACT-R model could also potentially be modified in order to predict learning trends in different task contexts and with new keypad layouts for which there is no existing empirical data. This research is still ongoing, though ACT-R modelling would seem to hold promise as a technique to be used as part of the product development process for mobile phones and other new devices.

4.4 Models of learning and problem solving in practice

In the previous three sections we have seen two complementary parts of the ACTR approach to learning. In Section 4.1 we considered how the ACT-R production compilation process can be used to create general rules from dependency goal chunks. In Section 4.2 we described how production rules, in this case a model of a counting procedure, can be used to create new facts in the form of chunks in declarative memory. In Section 4.3 we saw how the gradual retrieval of new chunks can be modelled and how learning may interact with visual and motor systems. So to what extent does ACT-R provide a general explanation of how humans learn?

Learning is modelled by ACT-R as an incremental process through which new knowledge is acquired declaratively, and performance gradually becomes faster and less error prone though practice. Such an account of learning would seem to accurately reflect how, for example, someone learns to change gears in a manual car. The instructor initially gives verbal instructions (i.e. a declarative account) on how to change gears. Initial attempts by the learner to change gear are slow, deliberate, error prone, and rely heavily on the declarative account. Months later the same person will be able to swiftly move through the gears, rarely making errors, and will no longer be using the declarative instructions of how to perform the process. An ACT-R account of this form of learning appears highly satisfactory. Performance gradually becoming faster, less error prone, and more automatic can be explained as being due to production compilation.

However, other forms of human learning are harder to explain, as learning may not always be characterized as a gradual increase in performance through experience. For example, children's learning of the past tense verbs in English doesn't quite follow this pattern (Bowerman, 1982). Initially children are very good at forming the past tense correctly. However, over time, and as they learn more words, their performance deteriorates, and they start tending to use the regular verb rule 'add –ed' even when they shouldn't, saying, for example, 'breaked' rather than 'broke'. Of course, over time, they overcome this problem, and the accuracy goes up again – forming a kind of 'U-shaped' pattern.

This pattern can be interpreted as being due to a progression through three ways of performing the task during learning. At first, each problem to be solved (e.g. each past tense verb to be constructed) is dealt with uniquely. The high degree of regularity (e.g. the large number of past tense words that can be created by just adding 'ed') is not reflected in the cognitive mechanism used to generate past tense verbs. Later, a single mechanism is used to solve an (overly) wide range of cases. The child now starts to use 'breaked' rather than 'broke'. Finally, exceptions are correctly handled and the child starts to say 'broke' once again.

A similar U-shaped pattern of performance over time has been found in children's ability to solve certain mathematical puzzles (Richards and Siegler, 1981) and in students' radiological diagnoses (Lesgold *et al.*, 1988). Lesgold *et al.* found that students roughly three or four years into a course perform worse than more experienced professionals, but also worse than they did the previous year. In terms of performance, it appears that sometimes a skill has to be acquired, lost, and then regained. Human learning therefore may be more complex than ACT-R might suggest.

SUMMARY OF SECTION 4

- Complex behaviour can be modelled using production rules, which contain a condition (IF part) and an action (THEN part).
- ACT-R has a production compilation mechanism for learning new production rules from instructions and examples.
- Learning can also involve the creation of new declarative chunks from production rules.
- Learning and problem solving behaviour in ACT-R, as driven by the production rules, is influenced by levels of activation.
- ACT-R tends to characterize learning as an incremental process through which performance gradually becomes faster and less error prone. Human learning does not always follow this pattern.

5 A COMPARISON OF ACT-R AND PDP

In the past few sections we have described the ACT-R cognitive architecture. Here we will compare ACT-R as an example of a rule-based cognitive architecture against the PDP cognitive architecture overviewed at the beginning of the chapter.

Children's learning of past tense verbs in English is a good point for comparison, as it has been modelled using both PDP models and rule-based models like ACT-R. The central question is: what is happening to cause the 'U-shaped' pattern effect? One possibility is that there is an area of memory that functions like a PDP network and that can generalize as it learns from examples. Rumelhart and McClelland proposed a model like this for learning past tenses.

In Rumelhart and McClelland's model, the changes in accuracy are caused by the child learning more new verbs. When there are only a few words, a network can fit in all the words by simply making strong links between individual input and output units. Each past tense verb is therefore dealt with uniquely. There comes a point when this won't work – when there are more words than available units – and then the network needs to build a more generalized association. For a short time the model over-generalizes, as a child does. This over-generalization is corrected in the PDP network through training (see Section 1.1), where the network receives feedback as to the correct answer.

Rumelhart and McClelland's two-layer model was relatively simple – it depended on this forced increase in vocabulary. Plunkett and Marchman (1991) used a three-layer network instead, and found that adding a hidden layer produced a similar 'U-shaped' pattern without needing to increase the vocabulary dramatically at one point in the learning process. They still needed a gradual increase in vocabulary to get the right pattern, however (Plunkett and Marchman, 1996).

In contrast, Taatgen and Anderson (2002) set out to model past tense learning using ACT-R, where there are two parts of memory involved: declarative memory and procedural memory. They suggested that children initially learn past tenses as declarative chunks, and then through a process of production compilation (see Section 4.1) form a slightly unreliable production rule to generate past tenses. Over time, they develop a blocking mechanism that stops the rule from being used when there is an exception stored in declarative memory. The dip in performance is caused by the unreliability of the production rule when initially constructed.

Both the PDP and ACT-R models correspond, more or less, to the observed behaviour, but the underlying explanations differ in a few subtle ways. For example, the PDP model depends on feedback. However, there is a problem with this explanation because children aren't always given feedback, and even when children are corrected, they still tend to use the over-generalized regular verb rule. Conversely, the dip in performance of the ACT-R model is due to the unreliability of the production rule when first formed. This unreliability is responsible for cases of over-generalization, but feedback is not required in order to correct this over-generalization. Taatgen and Anderson (2002) cite evidence that children also pick up correct past tenses from listening to others, largely without receiving feedback on their own performance.

The PDP and ACT-R models make different theoretical claims. The PDP model has the assumption that there is one area of memory, and change is driven by change in acquired vocabulary. The ACT-R model has the assumption that there are two areas of memory (declarative and procedural), and change is driven by practice. In principle, these differences can be tested, and experiments used to gather empirical evidence on the matter.

To sum up, the key difference between rule-based architectures, such as ACT-R, and the PDP approach is the nature of the representation used in the model. PDP models, being closer to Marr's hardware level, are described as having a sub-symbolic representation. **Sub-symbolic models** do not contain any explicit representations of symbols, such as production rules. Instead, they construct a sub-symbolic neural-like representation that can help support and explain a symbolic account of cognition. Conversely, rule-based systems explicitly use **symbolic representations**, such as the 'breaked' and 'broke', that might be found in the declarative chunks and production rules we have seen in this chapter.

In terms of their operation, PDP models, as their full name suggests, work in parallel, with signals being passed simultaneously throughout the network of artificial neurons. Symbolic cognitive architectures tend to be serial in operation. In ACT-R only a single production rule fires at any one time. The parallel processing of the network is mimicked in ACT-R by the activation equations that simultaneously update the activation of elements in declarative memory.

A final important difference between ACT-R and PDP concerns the kinds of cognitive phenomena that they most successfully explain. Sub-symbolic cognitive architectures such as PDP models, although able to emulate rules, are generally stronger at explaining automatic processes (e.g. face recognition). Symbolic cognitive architectures such as ACT-R are stronger at modelling consciously controlled processes such as problem solving.

As we have seen, ACT-R incorporates some features of the PDP approach. However, to what extent is it possible (or desirable?) to build a rule-based system fully integrated with a PDP architecture?

This was attempted by Lebiere and Anderson (1993) in the design of ACT-RN, in which aspects of the ACT-R architecture, such as chunks and the goal module, were implemented using a PDP network. Although successful to a degree, many features of ACT-R were difficult to model using PDP. This finding led them to actually remove many of these features from ACT-R, on the grounds that they were not neurologically plausible. For example, ACT-R used to allow a slot in a chunk to have a long list of items as its value. This cannot be done in the current version of ACT-R and lists must be represented as described in Section 3.1. The experiment with ACT-RN also led to PDP-like features being incorporated into ACT-R itself. Partial matching, as introduced in Section 3.3, is one such example. There is therefore reasonable justification to refer to ACT-R as a hybrid architecture that shows both PDP and rule-based characteristics.

However, it is still unclear to what extent it is possible to completely integrate symbolic and sub-symbolic architectures. And even if it is possible, it may not always be desirable. The models will necessarily be far more complex, and may inherit the weaknesses rather than the strengths of the symbolic and sub-symbolic approaches. This is one of the reasons why the current version of ACT-R has features motivated by PDP, such as activation, but does not encompass all features of PDP within it.

SUMMARY OF SECTION 5

- Rule-based architectures such as ACT-R provide a symbolic account of human cognition, operate in a largely serial way, and are particularly strong at modelling consciously controlled processes such as problem solving.
- PDP architectures provide a sub-symbolic account of human cognition, operate in parallel, and are particularly strong at explaining automatic processes such as face recognition.
- Attempts to completely integrate symbolic and sub-symbolic architectures have had limited success, but there appear to be advantages in the translation of certain coarse-grained features from one architecture to the other.

6 WHEN IS A MODEL A GOOD MODEL?

Modelling has a long and respectable heritage within psychology, with computers being used to model cognitive behaviour even at the birth of cognitive psychology itself, from 1956. Throughout, there has been a continuing question about how models fit in with experimental psychology. One very big question in cognitive modelling is: given a model, how do you know whether it is a good model or not? In this section we consider four criteria against which a model can be judged, followed by a description of the Newell Test, which constitutes an ambitious agenda for cognitive modelling. The four evaluation criteria we wish to consider are:

- The extent to which the behaviour of the model fits human performance.
- The validity of the model from the viewpoint of psychological theory.
- The parsimony of the model – the extent to which unnecessary complication is avoided.
- The biological plausibility of the model.

In Section 1, we defined cognitive modelling as building a model of a cognitive process and comparing the behaviour of the model against human performance. If the model behaves in the same way as humans, then the structure of the model and the way it works may give some insight into how humans perform the task. Clearly, if the behaviour of the model does not mirror human performance, then there is no support for the hypothesis that the internal workings of the model reflect human cognitive processes. And, of course, this failure of a model to fit the data can itself be an important and useful lesson learned.

However, it should not be assumed that the closer the fit to the empirical data, the better the model (Roberts and Pashler, 2000). Although a good model should at least roughly approximate to the data, the most closely fitting model is not necessarily the best. As described by Pitt and Myung (2002), a cognitive model could actually over-fit the data. A model may be so carefully customized to a specific set of empirical data that the generalizability of the model and its components to similar cognitive processes has been jeopardized. The extent to which the model fits the empirical data is therefore insufficient on its own as a measure of quality.

This leads us onto our second criterion. The internal structure of the model, by which it produces behaviour, needs to be defensible in terms of the psychological literature. As described in Section 1.3, one motivating factor in the development of cognitive architectures was to logically separate the cognitive model from the workings of the computer. The features of ACT-R available to the modeller, such as procedural and declarative memory, chunks, production rules, and production compilation, are the mechanisms by which the model produces its behaviour. Each of these features can be debated and compared against the psychological literature. This clear distinction between the model and its computer implementation has been one of the great successes of work into cognitive architectures.

Our third criterion is parsimony. The law of parsimony or **Ockham's Razor** states that an explanation of a phenomenon should not contain any unnecessary detail. Specifically, and in relation to cognitive modelling, a model should contain only the minimum number of components and so should not contain components that do not impact on the behaviour of the model. Therefore, any component of a model has to provide explanatory significance that justifies the additional complexity that it also brings. Ockham's Razor can also be used to criticize the careful fitting of a model to the empirical data, as this can increase the complexity of the model for little gain.

Finally, the biological plausibility of a model is important. To come back to the car analogy, if something looks like a car and moves from A to B on wheels, but you know there is a person inside pushing it along, there are limits to what it can tell you about how a car actually works. Both PDP and symbolic models have made simplifying assumptions about the nature of the brain, but an increasing attention to brain imaging is starting to make testable claims about how parts of the brain work together to achieve cognition, and these are increasingly important to cognitive modelling as a method (Anderson *et al.*, 2004).

Despite its increasing maturity, cognitive modelling still lacks a clear method for how models should be evaluated. However, the above criteria can be used to provide a broad evaluation of any cognitive model, whether it is a rule-based or a PDP model.

Other work in the area of model evaluation has aimed to devise and follow an ambitious set of criteria against which individual cognitive models and the progression of the cognitive modelling field as a whole can be tracked. Anderson and Lebiere (2003) elaborate Newell's (1990) 12 criteria for assessing the quality of a model, which they call the **Newell Test**. These are shown in Box 19.4.

Anderson and Lebiere (2003) gave ACT-R a grade for each point and, based on these criteria, there are some areas where ACT-R is strong. It is pretty good at exhibiting rational behaviour, at coping with error, learning, and at modelling real-time behaviour. But there are other areas where ACT-R is weaker, such as in using natural language, exhibiting self-awareness, and being realizable within the brain.

BOX 19.4 METHODS Constraints on a human cognitive architecture (after Anderson and Lebiere, 2003)

A successful model should:

1. Behave as an (almost) arbitrary function of the environment (universality)
2. Operate in real time
3. Exhibit rational (i.e. effective) adaptive behaviour
4. Use vast amounts of knowledge about the environment
5. Behave robustly in the face of error, the unexpected, and the unknown
6. Integrate diverse knowledge
7. Use (natural) language
8. Exhibit self-awareness and a sense of self
9. Learn from its environment
10. Acquire capabilities through development
11. Arise through evolution
12. Be realizable within the brain

However, these criteria should not only be used to highlight the strengths and weaknesses of cognitive architectures such as ACT-R and PDP, but also to show how researchers working with different architectures can learn from each other. Anderson and Lebiere claimed that the Newell Test could lead PDP researchers to incorporate ideas from ACT-R, similar to the way ACT-R has over recent years incorporated ideas from PDP. These criteria and the cognitive architectures they evaluate can therefore help to provide an overarching account.

As mentioned in Section 1.3, Newell (1973) argued that at that time cognitive psychology was asking lots of small questions, and progressing through small steps, but that the big picture was disjointed because there was little in the way of an overarching framework to glue the work together. He argued that a move to complete theories and models rather than partial ones, to complex composite tasks rather than narrow focused ones, and to models that would cope with many tasks rather than just one or two, would help the science of cognitive psychology to progress more effectively. In many senses, Newell's article laid the foundations for cognitive architectures like ACT-R and Newell's own Soar architecture (Newell, 1990). And the move to using cognitive models in conjunction with empirical studies and the development of cognitive theory to help connect diverse elements of cognitive psychology is set to continue.

As a set of points, though, Newell's list of criteria is helpful simply because it is so extensive. It shows just how far there is to travel before cognitive models are capable of explaining cognitive behaviour in an integrated manner. But we should not leave models on this note, as this issue and the list of criteria apply to all cognitive psychological theories, not just the kinds of model exemplified by ACT-R. ACT-R may still have a long way to go, but it is one of the best approaches available.

SUMMARY OF SECTION 6

- It is possible to set out many criteria to help judge the usefulness of a model.
- ACT-R performs reasonably well across the board, although it shows weaknesses in the areas of natural language, self-awareness, and biological plausibility.
- Models can be useful independently of the quality of their empirical predictions, in that they allow a community of researchers to be brought together to share ideas.

7 CONCLUSIONS

The advent of computers was central to the foundations of cognitive psychology. They provided both a new set of concepts that could be used to understand human behaviour and a new method that could be used to study it. Symbolic cognitive architectures such as ACT-R, however, are offering something more precise than computer metaphors of the human mind. Rather, they assume that the working of the mind is essentially the symbolic representation of knowledge (e.g. in the form of chunks) and the use and transformation of these symbolic representations in order to perform tasks (e.g. actions performed by production rules).

The ACT-R cognitive architecture can be used successfully to model a wide range of psychological phenomena such as recall effects in list memory and the development of skills in arithmetic and mobile phone texting. Each model makes use of, and provides a test of, the same underlying modelling constructs.

Symbolic cognitive architectures are emulated on a computer but can be thought of as distinct from the workings of the computer itself. The PDP cognitive architecture is also emulated on a computer, but here the assumption is that cognitive functions can be constructed from artificial neural elements having some similar properties to the human brain. The progressive development of ACT-R, and the incorporation of PDP-like properties into the ACT-R architecture, highlights a trend towards a hybrid approach to modelling that aims to combine the benefits of symbolic and sub-symbolic approaches.

Cognitive modelling involves not only building models but also evaluating them and the cognitive architectures that have been used to build them. A number of evaluation criteria have been proposed, encompassing issues such as the fit between the model and both human performance and psychological theory, its parsimony, and its biological plausibility.

FURTHER READING

Anderson, J.R. and Lebiere, C. (1998) *The Atomic Components of Thought*, Mahwah, NJ, Lawrence Erlbaum Associates. ACT-R has changed a little since this book was written, but it provides a good introduction to ACT-R and a selection of models that can be built using ACT-R, covering topics such as learning, memory, arithmetic, and analogy.

Anderson, J.R. and Lebiere, C. (2003) 'The Newell Test for a theory of mind', *Behavioural and Brain Sciences*, vol.26, pp.587–640. This paper describes Newell's 12 criteria as outlined in this chapter and ranks ACT-R against them. The paper is followed by an extensive commentary from a number of researchers in the field.

Anderson, J.R., Bothell, D., Byrne, M., Douglass, S., Lebiere, C., and Qin, Y. (2004) 'An integrated theory of the mind', *Psychological Review*, vol.111, pp.1036–60. This paper explains the updated ACT-R architecture as used in this chapter and illustrates its use through some example models.

REFERENCES

Anderson, J.R. (1974) 'Retrieval of prepositional information from long-term memory', *Cognitive Psychology*, vol.6, pp.451–74.

Anderson, J.R. (1990) *The Adaptive Character of Thought*, Hillsdale, NJ, Lawrence Erlbaum.

Anderson, J.R. and Bower, G.H. (1973) *Human Associative Memory*, Washington, Winston and Sons.

Anderson, J.R. and Lebiere, C. (1998) *The Atomic Components of Thought*, Mahwah, NJ, Lawrence Erlbaum Associates.

Anderson, J.R. and Lebiere, C. (2003) 'The Newell Test for a theory of mind', *Behavioural and Brain Sciences*, vol.26, pp.587–640.

Anderson, J.R., Bothell, D., Byrne, M., Douglass, S., Lebiere, C., and Qin, Y. (2004) 'An integrated theory of the mind', *Psychological Review*, vol.111, pp.1036–60.

Anderson, J.R., Bothell, D., Lebiere, C., and Matessa, M. (1998) 'An integrated theory of list memory', *Journal of Memory and Language*, vol.38, pp.341–80.

Baddeley, A. (1986) *Working Memory*, Oxford, Clarendon Press.

Bowerman, M. (1982) 'Reorganizational processes in lexical and syntactic development', in Wanner, E. and Gleitman, L.R. (eds.) *Language Acquisition, The State of the Art*, Cambridge, Cambridge University Press.

Butts, L. and Cockburn, A. (2002) 'An evaluation of mobile phone text input methods', *Australian Computer Science Communications*, 24, 55–9.

Chase, W.G. and Simon, H.A. (1973) 'Perception in chess', *Cognitive Psychology*, vol.4, pp.55–81.

Das, A. and Stuerzlinger, W. (2007) 'A cognitive simulation model for novice text entry on cell phone keypads', in *Proceedings of the European Conference on Cognitive Ergonomics: ECCE 2007*, London, UK.

Das, A. and Stuerzlinger, W. (2008) 'Modeling learning effects in mobile texting', in *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia: MUM 2008*, Umea, Sweden.

Hitch, G.J., Burgess, N., Towse, J.N., and Culpin, V. (1996) 'Temporal grouping effects in immediate recall: a working memory analysis', *Quarterly Journal of Experimental Psychology*, vol.694, pp.116–39.

Lebiere, C. and Anderson, J.R. (1993) 'A connectionist implementation of the ACT-R production system', in *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society*, Hillsdale, NJ, Erlbaum.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., and Wang, Y. (1988) 'Expertise in a complex skill: diagnosing X-ray pictures', in Chi, M.T.H., Glaser, R., and Farr, M.J. (eds.) *The Nature of Expertise*, Hillsdale, NJ, Erlbaum.

MacKenzie, I.S. and Zhang, S.X. (1999) 'The design and evaluation of a high-performance soft keyboard', *ACM Conference on Human Factors in Computing Systems (CHI '99)*, New York, ACM.

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, W.H. Freeman.

Newell, A. (1973) 'You can't play 20 questions with Nature and win: projective comments on the papers of this symposium', in Chase, W.G. (ed.) *Visual Information Processing*, pp.283–308, New York, Academic Press.

Newell, A. (1990) *Unified Theories of Cognition*, Cambridge, MA, Harvard University Press.

Norman, D.A. (1981) 'Categorization of action slips', *Psychological Review*, vol.88, pp.1–15.

Pitt, M.A. and Myung, I.J. (2002) 'When a good fit can be bad', *Trends in Cognitive Science*, vol.6, pp.421–5.

Plunkett, K. and Marchman, V. (1991) 'U-shaped learning and frequency effects in a multilayered perceptron: implications for child language acquisition', *Cognition*, vol.38, no.1, pp.43–102.

Plunkett, K. and Marchman, V. (1996) 'Learning from a connectionist model of the acquisition of the English past tense', *Cognition*, vol.61, no.3, pp.299–308.

Prince, A. and Smolensky, P. (1997) 'Optimality: from neural networks to universal grammar', *Science*, vol.275, no.14, pp.1604–10.

Rabinowitz, M. and Goldberg, N. (1995) 'Evaluating the structure-process hypothesis', in Weinert, F.E. and Schneider, W. (eds.) *Memory Performance and Competencies: Issues in Growth and Development*, Hillsdale, NJ, Lawrence Erlbaum.

Richards, D.D. and Siegler, R.S. (1981) 'U-shaped curves: it's not whether you're right or wrong, it's why', in Strauss, S. and Stavy, R. (eds.) *U-shaped Behavioural Growth*, New York, Academic Press.

Roberts, S. and Pashler, H. (2000) 'How persuasive is a good fit? A comment on theory testing', *Psychological Review*, vol.107, pp.358–67.

Rumelhart, D.E. and McClelland, J.L. (eds.) (1986) *Parallel-Distributed Processing: Explorations in the Microstructure of Cognition, Vol.1*, Cambridge, MA, MIT Press.

Siegler, R.S. (1988) 'Strategy choice procedures and the development of multiplication skill', *Journal of Experimental Psychology: General*, vol.117, pp.258–75.

Taatgen, N.A. and Anderson, J.R. (2002) 'Why do children learn to say "broke"? A model of learning the past tense without feedback', *Cognition*, vol.86, no.2, pp.123–55.

Trafton, J.G., Altmann, E.M., and Ratwani, R.M. (2011) 'A memory for goals model of sequence errors', *Cognitive Systems Research*, vol.12, pp.134–43.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

CHAPTER 20

THEORETICAL ISSUES IN COGNITIVE PSYCHOLOGY

Tony Stone and Martin Davies

1 INTRODUCTION

In the preceding chapters you have met a wide variety of theories of our cognitive capacities. In addition to debates over competing theories of specific capacities (e.g. of visual perception, autobiographical memory, or reasoning), cognitive psychology has seen debates that range more widely across the different areas of cognition. These concern the key concepts and explanatory strategies that are used in modelling cognition. This chapter introduces you to some of these debates. Before reading on, try Activity 20.1.

In doing Activity 20.1 you may have noticed that nearly all the cognitive theories discussed in this book make use of the notion of *mental representation* (in addition to the notion of *mental processing* mentioned in Activity 20.1). The Bruce and Young (1986) model of face recognition discussed in Chapter 4, for example, contains a component labelled *face recognition units* (FRUs): these are stored mental representations of the faces with which you are familiar. Similarly, models of spoken word recognition (Chapter 6) make use of the notion of a *mental lexicon* – this can be thought of as a store that contains mental representations of all the words that you know and representations of various properties of those words (e.g. their meanings). In the study of memory, there are thought to be mental representations of the episodes in one's life (in episodic memory) and of the 'know how' that one has (in procedural memory). In each of these examples, the idea is

✱ ACTIVITY 20.1

Look through the notes relating to theories or models of cognition that you have made on previous chapters and try to identify themes and concepts that recur.

COMMENT

You might start by simply making a list of all the theories or models that have been discussed in the previous chapters and then look for similarities and differences between them. If you look, for example, at the dual-route cascaded (DRC) model of word recognition (Chapter 6), you might note that there are two routes for the pronunciation of written words. One route involves assembling the phonological representation of the word using spelling-sound rules (also called the 'rule-based' route in Chapter 6), and the other route links the written form of the word with a stored representation of its correct pronunciation (also called the 'lexical' route in Chapter 6). You might then investigate the extent to which other cognitive models involve these two distinct kinds of *mental processing*.

that there is something internal to the mind that encodes information about the external world and about one's experiences, past, present, or future.

A fundamental principle of contemporary cognitive psychology is that cognition involves the storage and processing of information and that this information processing is achieved by *the processing or transformation of structured mental representations*. There are two basic kinds of processing in evidence in contemporary cognitive models. Sometimes representations are transformed by *rules*. For example, in Marr's model of the early stages of visual processing (see Chapter 3) we find the successive transformation of mental representations by a kind of rule called an algorithm. Thus, the grey level description is transformed into the raw primal sketch by an algorithm that compares images that have been blurred to different degrees. Similarly, in the DRC model of word reading, the pronunciation of non-words such as SLINT 'requires the postulation of a system of rules … loosely referred to as the letter-sound rule system' (Coltheart *et al.*, 1993, p.590). On other occasions information processing is conceptualized as the *transmission of activation* from one representation or group of representations to others. Whilst this kind of information processing is particularly characteristic of connectionist models of cognition (see Chapter 19), it is also found in more traditional models. In the Bruce and Young model of face recognition, there are connections between the FRUs and the person identity nodes (PINs) with activation flowing from an active FRU to the associated PIN.

The idea that cognition can be understood in terms of the rule-guided transformation of mental representations is at the heart of the computational model of the mind that has been dominant throughout most of the history of cognitive psychology. The first theoretical debate that we shall examine, in Section 2, concerns whether human information processing involves rules and representations to the extent proposed by many of the cognitive models you have studied.

A further thing you may have noticed when doing Activity 20.1, though you might have thought it too obvious to mention, is that there are different models for different cognitive capacities. Thus, there are separate models for spoken word recognition and visual word recognition (Chapter 6), for object recognition and face recognition (Chapter 4), for episodic memory and autobiographical memory (Chapters 8 and 17), and for speech comprehension and production (Chapters 6 and 7). This is not just a matter of practical convenience reflecting the specialized interests of different cognitive psychologists. Rather, it reflects the idea that the mind itself is composed of relatively independent, special-purpose systems that carry out specific information-processing tasks – that is, the idea that the mind is modular. Section 3 of this chapter is devoted to debates about **modularity**.

Debates on these two themes – rules and representations and the modularity of the mind – have been central to recent theoretical work, but there are other important questions that have also been subject to intense discussion. For instance, it is one of the attractions claimed for many connectionist models that they are more *brain-like* than traditional models. Yet it is noticeable that, in most of the preceding chapters, less is said about the way in which the various models of cognitive functioning are actually implemented in the brain. It may seem that there is only a poorly understood connection between, on the one hand, cognitive modelling and theories about cognitive processes and, on the other hand, the actual study of the physical brain. Nonetheless, much of the most compelling data for cognitive models over the past 35 years or so has come from the study of people who suffer from various kinds of cognitive impairment resulting from brain injury (Ellis and Young, 1998, give a textbook introduction to this work). To take just one example from many, the study of people with prosopagnosia (a specific inability to recognize once-familiar faces) has provided important evidence for the development of the Bruce and Young model of face recognition. This interaction between the study of normal and disordered cognition raises searching theoretical questions. For example, how exactly are inferences about normal cognitive processes made from cases of impaired cognitive functioning? These questions are taken up in Section 4.

The importance of neuropsychological evidence (Chapter 13) and the rapid development of neuroimaging techniques (Chapter 14) both highlight the need for a clear account to be given of the relationship between the kinds of models discussed in this book and the study of the brain. Will cognitive psychology ultimately reduce to – or will it perhaps be replaced by – neuroscience? These issues will be the topic of Section 5.

SUMMARY OF SECTION 1

- A fundamental principle of contemporary cognitive psychology is that cognition involves the processing of structured mental representations.
- Two kinds of processing are in evidence in cognitive models: rule-guided processing and the transmission of activation.
- Cognitive modelling reflects the idea that the mind is composed of modular systems.
- Inferences are made from neuropsychological case studies to models of normal cognitive functioning.
- There is a need to clarify the relationship between cognitive models and neurobiological models of the brain.

2 COMPUTATION AND COGNITION

One question that might occur to you as you near the end of this book is whether there is a *general* theory of the mind (or of mental functioning) that can be distilled from the preceding chapters. After all, one of the aims of science is to try to find general theories for seemingly disparate phenomena. One proposal for such a general theory has been influential throughout the history of cognitive psychology – indeed, it was present at its inception. This is the view that the mind is a computational device; that cognition is computation.

2.1 The computational model of the mind

We have already met the basic notions that lie behind this approach in discussing Activity 20.1 – it is the approach that sees cognition as the rule-guided transformation of structured mental representations. We refer to this idea as the **computational model of the mind (CMM)**.

An excellent example of this idea in action is the DRC model of word reading (Coltheart *et al.*, 1993, 2001) that you met in Chapter 6, and that was mentioned in Section 1. This model proposes that there is an assembled phonology route where the pronunciation of regular words and non-words is computed via spelling-sound rules. The pronunciation of the non-word SLINT, for instance, is arrived at through the application of the rules that S is pronounced /s/, L is pronounced /l/, I is pronounced /i/, and so on. It is important to be clear that these spelling-sound rules are not thought merely to be summary *descriptions* of the ways in which English regular words and non-words happen to be pronounced, but are meant to be part of the *causal* story of how people produce pronunciations of regular words and non-words of English. (Strictly speaking, familiar regular words will generally be pronounced via the lexical route, but they can also be pronounced via these spelling-sound rules.)

Let us compare the causal role played by spelling-sound rules in generating the pronunciations of words and non-words (according to the DRC model) with the role played by a rule that merely describes the operation of a system. Consider, for example, Ohm's law (this example is taken from Gallistel, 2001). Ohm's law states that in any electrical circuit, $I = \frac{V}{R}$ (where 'I' represents current, 'V' represents voltage, and 'R' represents resistance). The symbols 'I', 'V', and 'R' refer to measurable properties of an electrical circuit, and we can manipulate the symbols and make predictions that we can then test by measurement. For example, we can deduce from $I = \frac{V}{R}$ that $IR = V$, and then we can measure current and resistance, compute the numerical product $(I \times R)$, and see if the result matches what we find when we measure voltage (V). But the electrical circuit itself does not need to do the same computation. Ohm's law accurately *describes* the behaviour of electrical circuits, but computation guided by this law is not part of the *causal* story of why an electrical circuit behaves in the way it does. The electrical circuit itself does not contain *representations* of the current, voltage, and resistance, or a representation of the law that describes their relationship.

It is essential to the CMM that cognition is computation in the sense that it involves the *rule-guided processing* of structured mental representations. For processing to be *guided* by a rule, the rule must be part of the causal story of how the processing takes place.

Why is this approach called the *computational* model of the mind? The reason is that the digital computer is an example of a physical device that can process information by transforming symbols via a program – a set of rules stored in memory. Thus, the thought lying at the root of the computational view is that the brain (a physical device) processes information in the way that a computer does, in so far as human information processing involves transformation of mental representations and this transformation is guided by rules (akin to the computer program). It is important to recognize that this is a *model* of the mind. Models aim to capture what is fundamental to the thing being modelled – just as a geographical map (another kind of representation) attempts to capture what is fundamental to the terrain being mapped. So the CMM only aims to capture those aspects of the mind that are thought to be fundamental to information processing. Thus, the fact that a computer is made of silicon, wires, metal, and plastic (the computer hardware), whereas human information processors are made of flesh and blood (our hardware or, perhaps better, wetware), is not an objection to the CMM, because those aspects of a computer are not part of the model (and reasonably so given that the aim is to model the human mind, not human tissue).

But, you might wonder, shouldn't a model of the mind at least try to model that part of the human body – the brain – wherein mental processing occurs? Shouldn't we model the physical networks of neurons and the transmission of physical electro-chemical signals that actually implement information processing? It is a key aspect of the CMM that it does not try to model those neural processes as such. It does aim to provide a model of the brain, but at a level that is more abstract than the physical level. The CMM models 'the mind as the software of the brain' (Block, 1995). It is a matter of debate, of course, whether this degree of abstraction, this failure to model the brain's physical processes as such, is really a virtue, rather than a vice, of the CMM.[1]

2.2 Connectionist modelling

We turn now to the first of the theoretical debates that we want to consider in this chapter – connectionist objections to the CMM. Connectionism (also known as 'parallel distributed processing' ('PDP') and 'neural network modelling') is a style of psychological theorizing and model building that re-emerged in the middle of the 1980s, and has become massively influential in cognitive psychology since then.

Though there are many types of connectionist model, typically a model will be a network composed of three layers of artificial neurons or *units* – an input layer and an output layer with a layer of hidden units sandwiched between (Chapter 19). There are multiple *connections* between adjacent layers, and these connections transmit *activation* from input units to hidden units, and from the hidden units to output units. The level of activation passed from units in one layer to units in the next layer is a function of the activation of the units in the first layer (the units that are passing on activation) and the strength of the connections between these units and the units in the next layer (the units that are receiving activation). Through a process of adjusting the strengths of connections – the *weights* on the connections – such models are trained to associate patterns of activation across units in the input layer with patterns of activation across units in the output layer.

Patterns of activation in the input and output layers are the network's input and output *representations*. For example, written words of English might be represented by patterns of activation in the input layer and the output patterns of activation that the model is trained to produce might represent the pronunciations of those words. Networks of this kind have been developed as competitors to the DRC model of reading English words aloud (Seidenberg and McClelland, 1989; Plaut *et al.*, 1996). In the development of a connectionist model of reading aloud, the modeller has to decide how written words (made up of letters) will be represented by patterns of activation across the input layer. Will the representation of a word be made up of representations of letters? If so, will individual letters be represented by activation of single units, or by patterns of activation across several units? How will the order of the letters in a word be represented, so that BAT and TAB have distinct representations? Similar decisions have to be made about how pronunciations (made up of phonemes) will be represented by patterns of activation across the output layer.

A key aspect of connectionist models is that they can learn to associate patterns of input activation with patterns of output activation; they are not programmed in advance with an algorithm or rule that specifies what pattern should be associated with what. (There are various learning procedures, of which the most important is back-propagation of error. We need not go into the details here but see Chapter 19.)

At their most radical, connectionist modellers say that they aim to model human cognition in ways that dispense with rule-guided mental processing. For instance, McClelland and Patterson say that, in the connectionist approach to the psychology of language, 'cognitive processes are seen as graded, probabilistic, interactive, context-sensitive and domain-general. . . . Characterizations of performance as "rule-governed" are viewed as approximate descriptions of patterns of language use: no actual rules operate in the processing of language' (McClelland and Patterson, 2002b, p.465).

---

[1] It might be thought that the CMM must be wrong because we are conscious and computers are not. But this would be too quick. The CMM is a model and does not have to capture all aspects of human psychology. Perhaps a different model could deal with consciousness. However, it would be an objection to the CMM if failing to model consciousness meant that it thereby failed to model cognition – a claim made by John Searle (e.g. 1992).

Connectionism thus attacks a fundamental aspect of the CMM. It claims that mental processing is typically not rule-guided.

2.3 The past-tense debate: connectionism versus the CMM

We are going to present the connectionist challenge to rule-guided mental processing by considering the so-called past-tense debate. First, we set the scene for the challenge by describing an approach that does conform to the CMM.

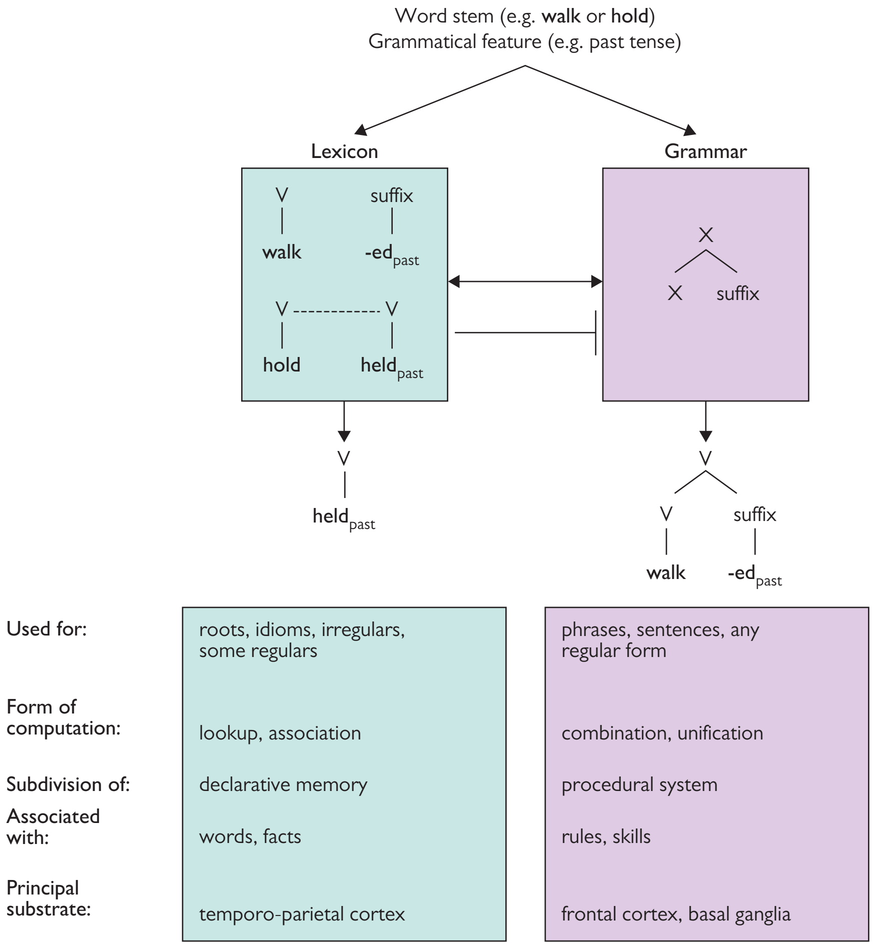
2.3.1 The words and rules model

The past-tense debate concerns how we should best understand the ability, possessed by all competent speakers of English, to form the past tense of English verbs. (The debate has also been concerned with how children *develop* the ability to form the past tense, but we leave such developmental questions to one side in this chapter.) This example is not as complex as some of the models and tasks you have met in this book, but it has been the subject of considerable debate, and the apparent simplicity of the phenomena being modelled allows one of the fundamental divides between the CMM and connectionism to be clearly seen.

Consider, then, the problem of how a competent speaker of English puts a verb into the past tense. The past tense of the overwhelming majority of English verbs takes the form VERB STEM + PAST TENSE MORPHEME. Thus the past tense of the verb TO HUNT is HUNT + ED (HUNTED), the past tense of the verb TO STROLL is STROLL + ED (STROLLED), the past tense of the verb TO JUMP is JUMP + ED (JUMPED), and so on. (Note that, although the written form of the past tense morpheme is the same in these examples, its phonology does vary.) However, there is a minority of verbs (around 180 of the most common in the English language) that do not follow this + ED pattern. These verbs form their past tense in seemingly irregular fashions. Thus the past tense of the verb TO GO is WENT (not GO + ED) and the past tense of the verb TO BUY is BOUGHT (not BUY + ED).

One way to model how people form the past tense has been developed by Steven Pinker and his colleagues (e.g. Pinker and Ullman, 2002a and b; Ullman *et al.*, 1997). This model – the **words and rules model** – 'claims that the regular-irregular distinction is an epiphenomenon of the design of the human language faculty' (Pinker and Ullman, 2002a, p.456). In outline, the model posits that two separate structures within the language faculty are responsible for the formation of the past tense: the lexicon and the grammar. When a past tense is to be formed, both the lexicon and the grammar are accessed in parallel. Verbs that have an irregular past tense will access the appropriate form in the lexicon. Verbs that have the regular past tense will access no past tense form in the lexicon, and the grammar component of the language faculty will therefore add the regular + ED ending or inflection. (This is, then, a *dual-route* model analogous to the DRC model of reading aloud, with which you are familiar.) Given that the lexicon and the grammar are accessed in parallel, the model posits an inhibitory signal from the lexicon to the grammar. Whenever an irregular past tense (e.g. BOUGHT) is accessed in the lexicon, this inhibitory signal then blocks the parallel rule-guided formation of an incorrect regular form (e.g. BUY + ED).

This model (see Figure 20.1) has been coupled with a hypothesis that draws on the distinction between declarative and procedural memory (Chapters 8 and 19). It has been proposed (Ullman, 2001; Pinker and Ullman, 2002a) that the lexicon is part of declarative memory and is subserved by temporal and temporoparietal regions of the neocortex, and that the grammar component is part of procedural memory and is subserved by the basal ganglia and the areas of frontal cortex to which they project. This proposal links the words and rules model with neural structures that are known, from independent evidence, to be involved in language processing. Moreover, independently of the debate about the past tense, declarative memory is thought to be responsible for the retention of facts,



	Lexicon	Grammar
Used for:	roots, idioms, irregulars, some regulars	phrases, sentences, any regular form
Form of computation:	lookup, association	combination, unification
Subdivision of:	declarative memory	procedural system
Associated with:	words, facts	rules, skills
Principal substrate:	temporo-parietal cortex	frontal cortex, basal ganglia

FIGURE 20.1 A simplified illustration of the words and rules theory and the declarative/procedural hypothesis. When a word must be inflected, the lexicon and grammar are accessed in parallel. If an inflected form for a verb (V) exists in memory, as with irregulars (e.g. *held*), it will be retrieved; a signal indicating a match blocks the operation of the grammatical suffixation process via an inhibitory link from lexicon to grammar, preventing the generation of *holded*. If no inflected form is matched, the grammatical processor concatenates the appropriate suffix with the stem, generating a regular form. *Source: Pinker and Ullman, 2002a, Figure 1, p.457*

and procedural memory responsible for the learning and control of motor skills, including skills that require sequencing. The thought, then, is that retention of an irregular past tense is the retention of a fact, whereas the construction of regular past tense forms depends on retention of a procedure.

2.3.2 Connectionist modelling of the past tense

Before reading on, try Activity 20.2.

Connectionist models aim to account for cognitive capacities without relying on rule-guided mental processing. How can this be achieved for the past tense? The first connectionist model of past-tense formation was developed by Rumelhart and McClelland (1986). This was a **pattern associator**, a simple network with just two layers of units (input and output units, but no hidden units) and one layer of connections. Patterns of activation across input units represent verb stems; patterns of activation across output units represent past-tense forms. (These are representations of the phonology, rather than the orthography (written form), of verbs and their past-tense forms.) The network was trained to associate the correct

✱ ACTIVITY 20.2

Try to identify the aspects of connectionist modelling that make it different from many of the other kinds of psychological models you have met in this book. For each difference try to find a specific connectionist model that illustrates the difference. In addition to this book, a good – freely available – web-based resource that will help you is the entry on connectionism (Garson, 2010) in the *Stanford Encyclopedia of Philosophy* (http://plato.stanford.edu/entries/connectionism/).

COMMENT

One difference you might find mentioned is that connectionist models are more 'brain-like' than models that are characteristic of the CMM. In what ways are connectionist models more 'brain-like' and are they very 'brain-like'?

past-tense form to both regular and irregular English verbs. Yet it did this with just a single route. All verbs, both regular and irregular, were processed via the same set of weighted connections from input units to output units. This appears to show that connectionist models are capable, in principle, of accounting for what looks like rule-guided behaviour in a way that does not require the use of a rule!

Rumelhart and McClelland's model learned to associate the appropriate past-tense form with each verb stem by relying on the statistical regularities contained in the training data. No rule was programmed into the model, and the trained model did not learn the task by representing and storing its own rule, such as 'add -ED to regulars'. How then does the model perform the task? Understanding this – and the same goes for any connectionist model – requires exploring in detail the activations of units and the weights on connections in the trained network. It should not be assumed, for instance, that the Rumelhart and McClelland model forms the past tense of all verbs in the way that the words and rules model forms the past tense of irregular verbs. It should not be assumed, that is, that the correct past-tense form of every verb is stored in a kind of declarative memory and then addressed by the appropriate verb stem.

The Rumelhart and McClelland (1986) model came in for some severe criticism from defenders of traditional rule-based theories (e.g. Pinker and Prince, 1988) but subsequent connectionist models have been able to meet some of those criticisms (e.g. Plunkett and Marchman, 1993; but see Pinker, 2006). What Rumelhart and McClelland (and other connectionist modellers) have shown, at least, is that there is an alternative to dual-route, 'words and rules' style models. But does the single-route connectionist alternative provide a better account of the available data than the dual-route words and rules model?

McClelland and Patterson (2002a) argue that connectionist models do provide a better fit with the data. They argue that the English past tense is a *quasi-regular* domain: in all but a few cases, past-tense forms that do not conform to the regular rule still display aspects of the regular pattern. Thus, McClelland and Patterson point out, the vast majority of English verbs with a so-called irregular past-tense form fall into one of nine groups, each of which exhibits aspects of the regular past tense. For instance, there is a group of verbs, including SAY, DO, TELL, and FLEE, whose past tenses are SAID, DID, TOLD, and FLED. Verbs in this cluster form the past tense by adding the /d/ sound as in the regular past tense but with a vowel adjustment to the stem. Another group, including BRING, CATCH, SEEK, TEACH, and THINK, form their past tenses by replacing the final consonant cluster with /t/ (the sound of the regular past tense following an unvoiced consonant, as in JUMPED) and adjusting the middle vowel to /O/ (sounds like 'aw'); thus BROUGHT, CAUGHT, and so on.

McClelland and Patterson claim that this quasi-regularity is not well explained by the words and rules model, in which formation of irregular and regular past-tense forms depends on two quite distinct mechanisms – the lexicon (words) and the grammar (rules). Quasi-regularity is, however, captured by connectionist models that have a single system for the formation of the past tense. McClelland and Patterson explain that their network can make the transformation KEEP → KEPT by simply:

> *adjust[ing] the activations of the output units representing the vowel, something the network will have learned to do on the basis of experience with* keep *and its neighbours* creep, leap, sleep, sweep, *and* weep. *The network uses the same connection-based knowledge that allows it to perform the regular mapping, and also taps into specific connections activated by the particular properties of keep to produce the vowel adjustment.*
>
> (McClelland and Patterson, 2002a, p.464)

In response, Pinker (2006) proposes a modified version of the words and rules model in which irregular past-tense forms are indeed stored in the lexicon, but the lexicon 'is not just a list of unrelated slots, but is partly associative'. Consequently, 'irregular verbs are predicted to show the kinds of associative effects that are well-modeled by pattern associators' (Pinker, 2006, p.223).

Pinker and Ullman (2002a; see also Pinker, 2006) draw attention to aspects of past-tense formation that they believe will be hard for connectionist models to deal with since their explanation lies in relatively deep linguistic principles. An example they provide is the way in which some usually irregular past-tense forms (e.g. STAND → STOOD) become regular in certain contexts such as GRANDSTANDED. These forms (regularizations of irregulars) occur due to linguistic principles concerning the formation of complex words of English. The verb TO GRANDSTAND is derived from a noun (A GRANDSTAND). Because of this, it is not possible for the irregular form (STOOD) stored in memory to be accessed – since that form must be accessed via a verb stem. Hence, no inhibitory block is sent from the lexicon to the grammar and the default rule is applied to form the regular past tense (GRANDSTAND + ED).

We shall return to the past-tense debate in Section 4.2.

2.4 Rules in connectionist networks

The leading idea of the CMM (the computational model of the mind) is that cognition involves the *rule-guided* processing of *structured* mental representations. If we are to understand connectionist modelling as a genuine alternative to the CMM, then we need to see how the information processing in connectionist networks departs from this idea. There is no doubt that *processing* takes place in networks and we have said that patterns of activation over input units, for example, are *representations* of items (such as words) in the network's task domain. In this section, we shall explore in what sense this processing might not be 'rule-guided' and, in the next (Section 2.5), in what sense the representations might not be 'structured'.

Rumelhart and McClelland draw attention to a difference between connectionist processing and rule-guided processing by saying that processing in a connectionist network does not require 'the postulation of explicit but inaccessible rules' (1986, p.218). According to the CMM, the rules that guide cognitive processes are often *inaccessible* to the person in whom those processes are taking place. We cannot simply introspect how cognitive processes work. (That is why we need to study cognitive psychology!) The CMM also allows that these rules, which guide cognitive processes, may be *explicit* in the sense that the rules are represented (in some format or syntax), are stored in memory, and have to be accessed before they can be used. A rule that is represented and stored, but is inaccessible to introspection, is sometimes said to be *tacitly known*. According to the CMM, there is a very clear distinction between having tacit knowledge of a rule and actually using that knowledge in processing. It is as clear as the distinction between having a cake recipe written in a book that is stored on the kitchen shelf and actually using the recipe to guide the processes of sieving flour, beating eggs, mixing ingredients, and so on, in order to make a cake.

In connectionist networks, rules are not explicit in this sense and there is a much closer relationship between having knowledge and using it. The processing in a network leads from patterns of activation across units in the input layer to patterns of activation across units in the output layer. The input patterns may represent written words and the output patterns may represent pronunciations; or the input patterns may represent verb stems and the output patterns may represent past-tense forms. If the network associates the correct pronunciations with written words or the correct past-tense forms with verb stems, this is because the weights on the connections have been adjusted (through a process of training such as back-propagation of error) to achieve this result. The trained network's knowledge of the task domain is thus embodied in the weights on the connections between units in adjacent layers. If the network's knowledge is in the weights then there is no need to access that knowledge before it can be used. When units in the input layer are activated, the activation is immediately passed forward from layer to layer, through the weighted connections, and a pattern of activation across the output units is produced. As McClelland *et al.* put it: 'Using knowledge in processing is no longer a matter of finding the relevant information in memory and bringing it to bear; it is part and parcel of the processing itself' (1986, p.32).

When we introduced the idea of rule-guided processing in the CMM, we said that rules are supposed to be part of the *causal* story of how a task is performed. For example, letter-sound rules are supposed to be part of the causal story of how the task of reading words aloud is performed. We contrasted this with the idea of rules as merely providing a summary *description* of patterns in the task domain. For example, it might be true of some domain of words that each word that begins with the letter B has a pronunciation that begins with the phoneme /b/. A successful model of the reading aloud task will perform in a way that *conforms* to this B → /b/ rule. But conformity to the rule does not tell us how the performance is actually achieved – it does not tell us whether the rule is merely *descriptive* or whether the rule is *causal*.

One possibility – in line with the CMM – is that the processing in the model is guided by explicit rules that are stored in memory, and have to be accessed before they can be used. An example of such a case is, again, the implementation of the assembled phonology route in the DRC model of reading. Coltheart *et al.* (1993) describe the model's use of spelling-sound rules as follows:

> *When confronted with a letter string for translation, it [the model] seeks to apply the rules to the string from left to right, starting with the longest possible rule that could accommodate that string. For the word* chip *it would start with four-letter rules, looking for a rule that maps the letters* chip *onto a single phoneme. No such rule will be found in the rule base. So a rule corresponding to the first three letters* chi *is sought; none will be found. The search for a rule for the first two letters* ch *[*ch → /tʃ/ *(sounds like '*ch*')] will, however, be successful.*
>
> (Coltheart et al. 1993, p.601)

You will not find explicit letter-sound rules in a connectionist model of reading aloud. But there is still a sense in which rules can figure in the causal story of how a connectionist model works and not just as a description of the network's performance. We can introduce this sense by considering a 'toy' connectionist network that takes representations of 25 written consonant–vowel pairs (built from five consonants and five vowels) and associates each of these with a representation of the corresponding pronunciation. For example, when given a representation of the written pair BA as input the network produces a representation of its pronunciation /bæ/ (sounds like 'ba' as in 'bat') as output. The network is represented in Figure 20.2.

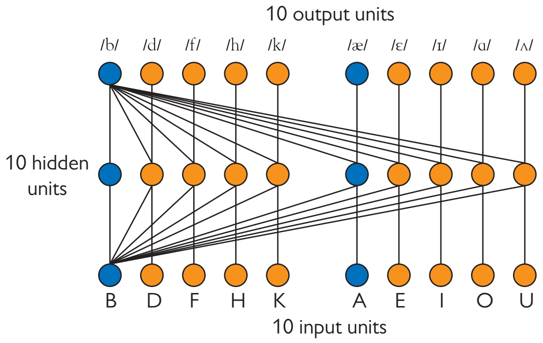


FIGURE 20.2 A connectionist model with implicit letter-sound rules (not all connections are shown).
Source: Based on Davies, 1995, Figure 3, p.181

This network associates input representations with output representations in a way that conforms to the B → /b/ letter-sound rule. How does the network achieve this performance? Examine the network shown in Figure 20.2 carefully. The input representation for a consonant–vowel pair such as BA involves activation of two units in the input layer (in this case, the consonant unit that represents B and the vowel unit that represents A). Whenever a pair beginning with B is presented to the system, the same input unit (representing B) is activated and, because the network conforms to the letter-sound rule, we know that the same output unit (representing the phoneme /b/) is activated.

Now, the important point to notice is that, whichever of the five pairs beginning with B is being processed, the causal explanation for the activation of the /b/ unit in the output layer is exactly the same. Activation is passed from the B unit to the leftmost hidden unit and on to the /b/ unit. So the five cases of conformity to the B → /b/ letter-sound rule have a *common causal explanation*. The explanation is grounded in a *causal common factor* – the passing of activation along the weighted connections from the B unit to the hidden unit and on to the /b/ unit. We could say that the B → /b/ rule is *implicit* in the network. The same, of course, goes for the nine other letter-sound rules that capture the pronunciations of these consonant–vowel pairs. This notion of a rule being implicit in a processing system requires more than just conformity to the rule. It requires that all the transitions that conform to the rule have a common causal explanation. But the notion of a rule being implicit in a system does not require that the rule is explicitly represented and stored and then accessed when it is needed.

Rules can figure in the causal story of how a connectionist model works because rules can be embodied implicitly in networks:

> *Is there, within the system [network], a component mechanism, or processor, or module that operates as* a causal common factor *to mediate all the input-output transitions that instantiate the pattern described by the rule? If so, then the rule is said to be implicit in the system (or the system is said to have implicit or tacit knowledge of the rule).*
>
> *(Davies, 1995, p.162)*

The point to take away from this discussion is *not* that all connectionist models embody implicit rules in this sense. There can certainly be networks whose performance conforms to rules, even though the input-output transitions that conform to a particular rule do not have a common causal explanation. When examining a connectionist model that performs a cognitive task, it is often useful to ask whether the performance is achieved by virtue of a structure of causal common factors that are the implicit analogues of explicit rules. If the network does not embody implicit rules, then we need to try to discover how the model does complete the task set for it.

2.5 Connectionism, structure, and compositionality

Connectionist modelling also challenges the style of *mental representation* that is characteristic of the CMM. This is a complex and (in parts) technical debate and we will briefly discuss just one issue.

The CMM holds that the mental representations processed by rules have **compositional structure** and that this structure explains a fundamental property of our cognitive systems – their **systematicity**. This latter property can most easily be introduced by reference to our linguistic abilities. Suppose I understand the sentences ‘The dog bit the cat’ and ‘The cow jumped over the moon’. My understanding of these sentences is systematic in that I need to learn nothing more in order to understand the additional sentences ‘The cow bit the dog’ and ‘The cat jumped over the moon’. This systematic understanding is possible because sentences are composed of words, and the same words, with the same meanings, can be deployed in different sentences to express new thoughts. It is this compositional structure of sentences that makes systematic understanding of a language possible.

Contrast this systematic understanding of our native language with what might happen when someone learns a second language from a phrase book. It is perfectly possible for a monolingual English speaker to learn to say the Romanian sentences ‘O sticle cu vin roşu’ and ‘Un caiine cu dinţ mici’ in appropriate circumstances and yet be unable to combine the words in these sentences in new ways to express appropriate new thoughts. Equally, an English speaker might learn to recognize and respond to utterances of those two sentences without being able to understand any other sentences of Romanian. You could achieve systematic understanding – going beyond those two sentences to use and understand additional sentences constructed from the same words – only if you knew what each word in the sentences meant and had some rudimentary grasp of the grammar of the language (how the words are put together). Sentences of Romanian do, of course, have compositional structure, just as sentences of English do. But knowledge of the meanings of a few phrases or sentences does not guarantee that you will become attuned to this structure (indeed, typically you will not). The CMM supposes, at least for the most part, that mental representations have compositional structure and that this structure is part of the causal story of how cognitive processes take place.

Connectionist modelling challenges this approach to mental representation. It is claimed, for example, that connectionist models can explain the systematicity of human cognition without appeal to mental representations with compositional structure (at least, in the sense of compositional structure used by the CMM).

Consider a simple example – this time in the domain of thought, rather than language. It seems to be a fact about human thought that if I am able to think the thought *New York is dangerous* and to think the thought *London is safe*, then I can also think the thought *New York is safe* and the thought *London is dangerous* (systematicity again). The capacity to think the first two thoughts seems to be intrinsically linked to the capacity to think the second two thoughts. The CMM explains the systematicity of thought by saying that the mental representations that underpin thinking have compositional structure, just as the sentences of a language have compositional structure. This is the *language of thought* hypothesis (Fodor, 1975, 1987; see also 2008).

It is certainly possible for connectionist representations to be compositionally structured. The network in Figure 20.2 represents the written consonant–vowel pair BA by activation of two units in the input layer, a consonant unit that represents B and a vowel unit that represents A. The consonant–vowel pair DE is represented by activation of a consonant unit that represents D and a vowel unit that represents E. Recombination of the resources in these two representations allows the network to represent the pairs BE and DA. Compositional structure allows systematicity of representation. In a similar way, a network might represent *New York is dangerous* by activation of a unit that represents the city of New York and another unit that represents danger. The thought *London is safe* might be represented by activation of a unit that represents the city of London and a unit that represents safety. Recombination of these representational resources would allow the network to represent *New York is safe* or to represent *London is dangerous*. Once again, compositional structure explains systematicity, just as it does according to the CMM.

Although connectionist modelling of cognitive processes allows for the possibility of representations with compositional structure, it does not require compositionality. There might be no pattern of activation across units that is present in every representation of a consonant–vowel pair beginning with B. There might be no pattern of activation that is shared by all of a network’s representations of thoughts about New York. In a limiting case, each consonant–vowel pair, or each thought, might be represented by activation of a distinct unit or pool of units.

It is at this point that the challenge to the CMM becomes clear because, even without compositional structure, systematicity can, in a way, still be preserved. The modeller just needs to make sure that, if there is an input unit that represents BA and another that represents DE, then there are two additional units that represent BE and DA. If there are separate pools of units that represent the thought *New York is dangerous* and the thought *London is safe* then there need to be two additional pools of units to represent the thoughts *New York is safe* and *London is dangerous*. If these conditions are met, then there can, indeed, be a kind of systematicity without compositional structure. But does this approach promise an illuminating account of the systematicity of language understanding or the systematicity of thought?

An advocate of the CMM may argue that this connectionist way of delivering systematicity without compositional structure fails to capture the idea that the capacity to think *New York is dangerous* and *London is safe* is intrinsically linked to the capacity to think *New York is safe* and *London is dangerous*. When the connectionist approach departs from the CMM, it makes systematicity look like an accident. A modeller who adds two separate pools of units to represent the thoughts *New York is safe* and *London is dangerous* could just as well – for all that connectionist representation requires – have added two separate pools of units to represent the thoughts *Life is short* and *Prunes go well with custard*. Connectionists are not without reply, of course, but we shall not pursue the issue further here.

SUMMARY OF SECTION 2

- The computational model of the mind (CMM) is the view that cognition is computation.
- The CMM sees mental processing as involving the rule-guided transformation of structured mental representations.
- Connectionist models are computational models that challenge the idea that mental processing is rule-guided. In connectionist models the primary role is played by the transmission of activation.
- In the past-tense debate, the dual-route words and rules model illustrates the CMM approach. Connectionist modellers have demonstrated that a single-route model can be trained to form the past tense of regular and irregular verbs by transmission of activation.
- Connectionist models do not make use of explicit rules and a network's knowledge of the task domain is embodied in the weights on connections. Nevertheless, there is a sense in which rules can figure in the causal story of how a connectionist model works.
- Connectionist models also challenge the need for mental representations to have compositional structure.

3 MODULARITY

Modularity should have been near the top of the list of concepts and themes relating to theories or models of cognition that you identified in Activity 20.1. But exactly what is meant when a psychological system or function is said to be modular? A notion of modularity is familiar from everyday life. Stereo systems and kitchens have modular components: a CD player connected to an amplifier and a pair of speakers; a 600 mm drawer unit next to a 1200 mm cupboard unit. The electrical wiring in a home is also modular, with one circuit for the lights, another for the sockets, and a separate circuit for kitchen appliances.

In cognitive psychology, it is important that we apply this everyday notion of modular components at the appropriate level, the level of information processing. The brain has anatomical components, but we cannot simply assume that modular information-processing components map directly onto anatomical components of the brain. Cognitive modules are distinguished from each other *functionally*; that is, not by what they *are* neurally, but by what they *do* computationally.

One aspect of the everyday notion of modularity is that modular components are somewhat independent from each other. It is possible to upgrade or even replace an amplifier, while continuing to use the same CD player and speakers, or to replace the CD player with a turntable. Similarly, in the kitchen, the cupboard unit can continue in use even if the drawer unit is removed to make way for a dishwasher. When there is an electrical problem in the home, one circuit may need its fuse replaced (or its circuit breaker reset) while the other circuits continue to operate normally.

David Marr argued that, without modularity, computational processes would be hard to debug or upgrade and he proposed a *principle of modular design* for computational systems: 'a large system can be split up and implemented as a collection of parts that are as nearly independent of one another as the overall task allows' (Marr, 1982, p.102). In a similar spirit, Tim Shallice speaks of modules as 'isolable subsystems' and endorses a definition of modularity proposed by Endel Tulving, who suggested that two psychological systems are functionally different if the following conditions hold:

> *One system can operate independently of the other although not necessarily as efficiently as it could with the support of the other intact system. The operations of one system could be enhanced without a similar effect on the operation of the other; similarly the operations of one system could be suppressed without a comparable effect on the activity of the other. The functional difference also implies that in important, or at least non-negligible ways, the systems operate differently, that is, that their function is governed at least partially by different principles.*
>
> (Tulving, quoted in Shallice, 1988, p.21)

In terms of the everyday notion of modularity, it is relatively uncontroversial that the mind is modular: cognition depends on information-processing systems that have somewhat independent components. These modular components are displayed in cognitive psychologists' box-and-arrow diagrams. **Q1**

Jerry Fodor (1983) has, however, provided us with a more detailed – and more controversial – discussion of modularity in his book *The Modularity of Mind*. This will be the topic for the remainder of this section.

3.1 An outline of Fodor's theory of modularity

According to Fodor's account, the mind is divided into three different types of system: (1) sensory transducers; (2) modular input systems; and (3) non-modular central systems (see Figure 20.3).

The sensory transducers pick up physical stimuli from the environment – photons hitting the retina, sound waves causing the tympanic membrane to vibrate, etc. – and transform these stimuli into a format or code that the brain can understand. In short, transducers produce internal representations in response to stimuli from the external world. Recall that in Chapter 6 (Figure 6.1) you were provided with a picture of the waveform of a segment of speech. That picture is a representation of some of the physical properties of the speech stream. The sensory transducers have to transform those physical properties into a format that the language-processing system can understand. Once external stimuli

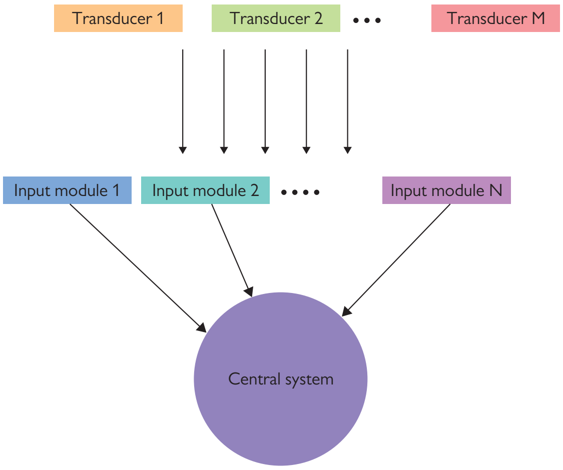


FIGURE 20.3 Fodor's model of perception and cognition. *Source: Coltheart, 1999, Box 1, p.116*

have been internally represented, computation can begin.

The modular input systems mediate between the sensory transducers and the central systems. They deliver, we might say, the perceptual experiences that provide a database of evidence for the processes of belief formation and decision making that are jobs for the non-modular central systems. The non-modular central systems also contain all the encyclopaedic knowledge one has stored in memory. In forming beliefs and making decisions, the central systems will typically take account of evidence from perception and also stored knowledge.

Fodor characterizes the modular input systems as those that possess the following cluster of properties:

- Mandatory operation
- Fast speed of operation
- Fixed neural architecture
- Characteristic and specific patterns of breakdown
- Characteristic pace and sequencing of their ontogeny
- Shallow outputs
- Limited central access to the mental representations that modules compute
- Domain specificity
- Informational encapsulation

This list of nine properties clearly goes well beyond the idea we introduced earlier of a module as a functionally somewhat independent component of a larger information-processing system.

The first five properties on the list are fairly straightforward. When Fodor says (the sixth property) that the final outputs of modules, which are available to central systems, are 'shallow' (1983, p.86), he means that the interface between modular input systems and central systems comes relatively early. For example, when you hear an utterance, the identity of the words and the grammatical structure of the sentence might be computed by a modular input system. But discerning the full significance of the utterance (working out, for example, whether the speaker was asking a question or making a suggestion, being ironic, or intending the utterance to be interpreted metaphorically), and deciding how to respond to the utterance, are tasks for the non-modular central systems. The seventh property of modules – that there is only limited central access to the mental representations that they compute – is the idea that the information at the various intermediate stages of a module's computation is not generally available to central systems.

Fodor does not say that a psychological system must possess *all* of the nine properties on the list if it is to be modular; indeed, he explicitly states that his concept of modularity is a cluster concept and that modularity admits of degrees. However, it is clear from Fodor's other writings that some of these properties – in particular, **domain specificity** and especially **informational encapsulation** – are more important to his conceptualization of a modular input system than others (e.g. Fodor, 1985).

3.1.1 Domain specificity

In *The Modularity of Mind*, Fodor's main examples of modular input systems are those responsible for visual perception and for the recognition of spoken language. The claim that these systems are *domain specific* amounts to the idea that these input systems deal only with a limited, idiosyncratic range of stimuli. The language module, for example, only deals with linguistic input. Thus it processes spoken language, but not more general environmental sounds such as the sound of a bell tolling. You should note that, because the language module can also process written language and sign language, the modular input systems do not correspond in a one-to-one way with the five senses.

In cognitive psychology, it is domain specificity that is most often associated with claims about modularity, although the modules proposed are often more fine-grained than those discussed by Fodor. Cognitive psychologists have proposed, for instance, that the visual system includes three modular sub-systems that have the specific jobs of processing faces, processing objects, and processing written words (see Chapters 4 and 6). Moreover, in cognitive psychological models, modules can be nested within one another – for instance, the face-processing module might itself be broken down into sub-modules. Finer-grained domain specificity and modules within modules can be taken as wholly within the spirit of Fodor's account.

3.1.2 Informational encapsulation

Informational encapsulation is the property that is non-negotiable for Fodor as regards whether a system is granted the status of a module. He describes the informational encapsulation of input systems as 'the essence of their modularity' (1983, p.71). We need, therefore, to be clear about what this property involves.

The easiest way to do this is via an example. Look at Figure 20.4. Do the two horizontal lines look the same length? It would be surprising if they did. For most viewers, the top horizontal line *looks* longer than the bottom one. The two lines are, in fact, of equal length (measure them if you want to check). With this knowledge in mind, look at the picture again. Do the two lines *now* look the same length? Once again, it would be surprising if they did. You still see exactly what you did before. But, because you know that the two lines are really the same length, you don't believe what you see. This is an example of *informational encapsulation*. You (now) have some knowledge – about the lines in the Ponzo illusion being the same length – but that knowledge does not affect the processing that takes place within the visual-input system. The visual-input system is, if you like, sealed off from that information or, as Fodor also describes it, visual perception is *cognitively impenetrable* – thoughts and beliefs are unable to penetrate the modular visual input system. The general point – well illustrated by visual illusions such as the Ponzo illusion – is that the processing in a modular input system is not affected by information stored in the central systems.

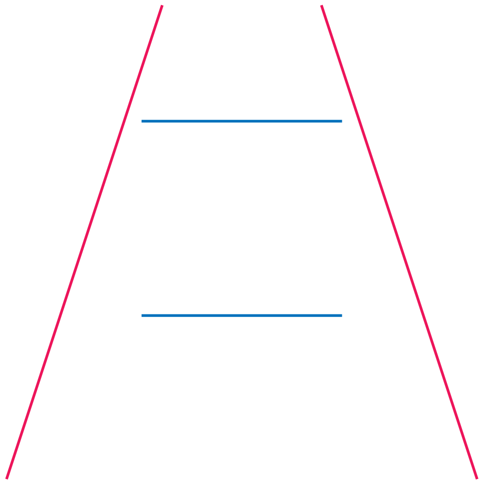


FIGURE 20.4 The Ponzo illusion.

In thinking about modules and the property of informational encapsulation, it is important to observe that Fodor does not rule out top-down information flow *within* a module. The top-down information flow that is excluded or restricted by informational encapsulation is the flow from central systems to modular input systems. The processing in a module does not draw freely on all the information that is stored in the central systems. This point about informational encapsulation has been a source of some confusion about Fodor's views on modularity, so it is worth spending a little more time discussing it.

In *The Modularity of Mind*, Fodor considers an objection to the idea that language comprehension is informationally encapsulated, an objection that stems from our ability to select the relevant meaning of an ambiguous word when it occurs in an appropriate context, such as 'Because Bruce was running short of cash, he went to the *bank*'. One possible explanation of this ability is that contextually relevant knowledge stored in the central systems influences processing in the language-input system, specifically, the process of lexical access. Thus, on the basis of your understanding of the early part of the sentence, your general knowledge about financial matters ensures that only the relevant meaning of the word BANK is accessed; or, more accurately, that only one of the two words spelled BANK is accessed in the lexicon.

Fodor draws on the results of Swinney (1979), which you have already met in Chapter 6, to question this account of contextual disambiguation. We will return to Swinney later, but for now we focus on one aspect of his findings. Swinney used a cross-modal priming paradigm in which participants listened to sentences, some of which contained an ambiguous word such as BUGS. Immediately following this word, a visual word (or non-word) was presented on a screen and the participants' task was simply to make a lexical-decision response ('Is this a word or not?'). In one condition, the ambiguous word BUGS was presented in the context of a sentence that did not bias interpretation towards one or the other of the word's meanings. In this condition, recognition of both the word ANT (semantically related to BUGS = insects) and the word SPY (semantically related to BUGS = secret radio transmitters used in espionage) was facilitated (by comparison with an unrelated word). This finding supports the hypothesis that, when there is no contextual bias, both meanings of BUGS (more accurately, both words with that spelling) are accessed.

In a second condition, the heard sentence strongly biased interpretation of the ambiguous word towards one of its meanings rather than the other. Swinney found that, when BUGS was presented in the context of a sentence about insects, recognition of the word ANT *and* the word SPY was once again facilitated. This finding supports the hypothesis that, when BUGS is first presented, both meanings (more accurately, both words with that spelling) are accessed despite the contextual bias, and that the locus of contextual disambiguation is *after* lexical access.

Fodor suggests that Swinney's findings can be accounted for in terms of the organization of the mental lexicon itself. Perhaps there are links between the lexical entries for semantically related words such that presentation of a word leads to the transmission of activation to lexical entries for semantically related words. In the case of an ambiguous word (e.g. BUGS), there is transmission of activation to lexical entries for words related to both of its meanings (ANT and SPY). Since the lexicon is *internal* to the language-input system, the semantic priming effects would not provide evidence that information from the central systems influences processing in the language-input system. As Fodor warns, 'it makes a difference … where the information comes from' (Fodor, 1983, p.80). If the information does not come from outside the module, it fails to be a counter-example to encapsulation.

✱ ACTIVITY 20.3

Fodor divides the mind into modular input systems and non-modular central systems. Which of the following do you think are modular input systems and which are non-modular central systems? In each case give reasons for your choice.

- Object recognition
- Face recognition
- Spoken word recognition
- Autobiographical memory
- Encoding and retrieval systems
- Problem solving
- Reasoning
- Spoken language production
- Reading aloud
- Attention
- Working memory

COMMENT

This is not as straightforward as it might seem. The earlier stages of object recognition and face recognition certainly seem to be input systems. They are modular in that they are domain specific, but do you agree with Fodor that they are informationally encapsulated? Problem solving seems to be a good candidate for a non-modular central system. However, many would argue that memory is modular in that there seem to be different systems for different kinds of memory (declarative/procedural/autobiographical) – but memory isn't an input system. How should attention be classified? Is it an input system? Is it modular? And what is to be said about output systems? Fodor didn't talk about output systems in *The Modularity of Mind*, but they are often taken to be modular. Do you agree? Consider, for example, one output function – reading a word aloud: which of the criteria for modularity does it meet?

3.2 The central systems

The central systems store information corresponding to knowledge and long-term beliefs and use this information in processes of reasoning, decision making, and planning. According to Fodor, there is an asymmetric flow of information between the modular input systems and the central systems. The input systems pass information – the outputs of modules, the results of their computational processes – to the central systems. But the processing in input systems is encapsulated from information stored in the central systems. This asymmetric flow of information brings clear benefits. If the processing in input systems were to be overly influenced by information already in the central systems – what we already know and believe – then we could never learn anything new from experience.

When we learn from experience, new information from input systems is accepted and old information stored in the central systems has to be revised. But there is a balance to be struck between new information and old information. It would be a very unwise creature that always accepted new information (or misinformation) delivered by an input system and ignored or discarded things that it already knew and believed. You should not always believe your eyes! The balanced integration of new information delivered by input systems with previously stored knowledge and beliefs is a crucial function of the central systems. It leads to the formation of new beliefs and the revision of old ones, and these new or revised beliefs are then used in reasoning, decision making, and planning.

Fodor makes two important claims about properties of the central systems. First, central systems are *domain general* rather than domain specific (they must be able to cope with the outputs from many different input systems). The reason is that it would be very unwise to update beliefs on the basis of the output from just one input system while other input systems were delivering different, and perhaps conflicting, information. Second, central systems are *unencapsulated* – in settling on what to believe, or in making a decision about what to do, anything and everything that the organism knows might, in principle, be relevant.

3.2.1 Argument from analogy: belief formation and hypothesis confirmation

Fodor's argument for the claim that the central systems are unencapsulated depends on an analogy between the processes in the central systems that lead to the *formation of beliefs* and the type of **non-demonstrative inference** that is characteristic of the *confirmation of hypotheses* in science. This is **abductive inference** or inference to the best explanation.

Scientific confirmation, Fodor says, is characterized by two properties that he dubs **isotropic** and **Quinean** (the latter after the famous twentieth-century American philosopher Willard van Orman Quine, 1908–2000, who argued that our knowledge was organized into a holistic web of belief). Fodor explains these notions as follows:

> *By saying that confirmation is isotropic, I mean that the facts relevant to confirmation of a scientific hypothesis may be drawn from anywhere in the field of previously established empirical … truths. Crudely: everything that the scientist knows is, in principle, relevant to determining what else he ought to believe. In principle, our botany constrains our astronomy, if only we could think of ways to make them connect.*
>
> *By saying that scientific confirmation is Quinean, I mean that the degree of confirmation assigned to a given hypothesis is sensitive to the entire belief system; as it were, the shape of our whole science bears on the epistemic status of each scientific hypothesis.*
>
> (Fodor, 1983, p.105)

These two properties, and the difference between them, need a little spelling out.

Suppose that you are considering two competing empirical hypotheses – perhaps two hypotheses about the generation of past-tense forms of regular and irregular verbs. As a cognitive psychologist, you can devise some behavioural experiments, expecting that the results may support one hypothesis or else the other. But you cannot claim that the evidence from your experiments is the only evidence that can be relevant to the choice between the competing hypotheses. Nor can you claim that the *kind* of evidence yielded by your experiments (perhaps reaction times for a task of producing the past-tense form of a verb) is the only kind of evidence that can be relevant. There is simply no way of putting a boundary around the kinds of evidence that may be relevant to the confirmation or disconfirmation of a hypothesis. As Fodor says of a closely related example: '*Any* facts about the use of language, or about how it is learned, or about the neurology of speaker/hearers, or, for that matter, about the weather on Mars, could, in principle, be relevant to the choice between competing linguistic theories' (1981, p.199). This is what Fodor means when he says that confirmation of hypotheses in science is isotropic. (Fodor also points out (1983, pp.106–7) that the processes of scientific *discovery* – the formation, rather than confirmation, of hypotheses and theories – seem to depend on analogical reasoning and are even more clearly isotropic.)

Saying that confirmation is Quinean is different from saying that confirmation is isotropic, even though being isotropic may sound rather similar to being 'sensitive to the entire belief system'. Suppose that we have two competing empirical hypotheses, or two competing theories, that fit all the relevant evidence equally well. Then we may still choose one theory over the other because it provides a *better explanation*. One theory may have the explanatory virtue of being simpler, or a better fit with the rest of science, or a less radical departure from previous theories (more conservative). In fact, whenever we evaluate competing scientific theories, we consider the explanatoriness of the theories as well as their probability given the available evidence. Fodor points out that, when theories are assessed for their explanatoriness, they are considered, not in isolation, but against the background of the rest of science. This is particularly clear in the case of the explanatory virtue of offering a good fit with the rest of science (offering good prospects for unification of phenomena that are superficially disparate). But the same point holds for the explanatory virtues of simplicity and conservativeness as well. Thus, the choice between competing hypotheses, or competing theories, depends on 'the shape of our whole science' (Fodor, 1983, p.105). This is what Fodor means when he says that confirmation of hypotheses in science is Quinean.

If Fodor is right to propose that the confirmation of hypotheses in science is a model for the formation of beliefs in the central systems, then we should expect that the operation of the central systems is isotropic and Quinean. An example illustrating what it means for central system processes of belief formation to be isotropic and Quinean is provided in Box 20.1. If the operation of the central systems *is* isotropic and Quinean then the central systems are not modular and, in particular, are not encapsulated. Processes of belief formation need to have free access to anything and everything that we know and believe.

This is Fodor's argument from analogy, but is there any direct empirical evidence that the central systems are unencapsulated? Is scientific confirmation really a good model of ordinary belief formation? Unfortunately, according to Fodor, evidence on this is scanty 'given the underdeveloped state of psychological theories of thought and problem solving' (Fodor, 1983, p.112).

3.2.2 Unencapsulated central systems and the CMM

As by his lights evidence is lacking, Fodor points to a difficulty in the development of theories of the central systems that is 'just the sort [of problem] we should expect to encounter if such processes are, in essential respects, Quinean/isotropic rather than encapsulated' (1983, p.112). It concerns the difficulties posed for artificial intelligence by the so-called frame problem. This is, very roughly, the problem of how to build a robot that can update its beliefs about the world as a result of the action it takes (Dennett, 1984). Consider, for example, a robot that is given the job of making a phone call to Mary:

> *Let's assume the robot 'knows' it can get Mary's number by consulting the directory. So it looks up Mary's number and proceeds to dial. So far, so good. But now, notice that commencing to dial has all sorts of direct and indirect effects on the state of the world (including, of course, the internal state of the robot), and some of these effects are ones the device needs to keep in mind for the guidance of its future actions*

BOX 20.1 Insects or espionage?

Suppose that you overhear a man in a dark suit utter an isolated sentence in which the word BUGS occurs. Perhaps he is talking about insects, or perhaps about secret radio transmitters. Suppose that, for some reason, it is crucial that you should reach the correct belief as to whether the man is talking about insects or espionage. Then, anything and everything that you know might be relevant.

First, *any* piece of evidence might turn out to be relevant. Does the man look as if he works for a pest control company, or does he look as if he might be working for (or against) an intelligence agency? Was there something in the newspaper about an invasion of insect pests, or about a spy cell being active? What about that report on the web about an interdisciplinary team of astronomers and biologists who claimed to have discovered a correlation – as yet unexplained – between the weather on Mars and the reproduction of ants? If there is a correlation, then the weather on Mars is relevant, just as Fodor said. But what about that other team who said they had discovered a different correlation between the weather on Mars and increased international tensions? Which team of scientists is to be trusted? Which universities are they from, and how good is the research there? So university research rankings are relevant as well. And so on, without end. If you really need to know whether it is more likely that the dark-suited man was talking about insects or about espionage, there is no good way of putting a boundary around the potentially relevant considerations and ignoring the rest of what you know. Belief formation is isotropic.

Second, the overall shape of your *entire* body of knowledge is important as well. Which hypothesis would fit better with what you already know and believe? How much revision to your other beliefs would be needed if you were to adopt the belief that the man was talking about insects because there is an infestation of ants, or if you were to believe that he was talking about espionage because there are secret radio transmitters everywhere? Which hypothesis would, perhaps after some initial upheaval in your other beliefs, allow the simpler and more coherent total world view? Belief formation is Quinean; it is 'sensitive to global properties of belief systems' (Fodor, 1983, p.111).

> *and expectations. For example, when the dialing commences, the phone ceases to be free to outside calls; the robot's fingers (or whatever) undergo appropriate alterations of spatial location; the dial tone cuts off and gets replaced by beeps; . . . and so forth. Some (but, in principle, not all) such consequences are ones the robot must be designed to monitor since they are relevant to 'updating' beliefs upon which it may come to act. Well, which consequences?*
>
> (Fodor, 1983, p.113)

We, of course, do this everyday updating effortlessly. But, at least on Fodor's reading of the literature, AI researchers have found building a device that can do it intractable.

It may be possible for AI researchers to solve or sidestep the frame problem, either by adopting ambitions more modest than building a robot with human intelligence, or by abandoning the computational paradigm that is the basis of the CMM. But Fodor regards the frame problem in AI as illustrating a much more

✱ ACTIVITY 20.4

Reconsider Chapters 10 and 12 on problem solving and on reasoning. Is there any evidence provided in those chapters that would either support or undermine Fodor's views on the central systems? Is he right that psychological theories of thought and problem solving are underdeveloped?

general problem about the processes of updating or revising beliefs. The problem arises if the processes involve updating in the light of the consequences of one's own actions (as in the example of the robot making a phone call). But it also arises if the updating is based on new information of other kinds (delivered, for example, by a perceptual input system). The processes of belief formation and belief revision need to be isotropic. There is no *a priori* method for putting a boundary around the considerations that might be relevant. And, as in AI, so more generally, this property of processing in the central systems presents a challenge to computational theories and, specifically, a challenge to the CMM (Shanahan, 2009).

Fodor believes that the prospects for a computational psychology of the central systems are bleak. Indeed, he proposes a principle that, he suggests, 'some day will come to be known as "Fodor's First Law of the Nonexistence of Cognitive Science"'; namely, that 'the more global (e.g. the more isotropic) a cognitive process is, the less anybody understands it. *Very* global processes, like analogical reasoning, aren't understood at all' (Fodor, 1983, p.107). According to Fodor, the CMM just will not explain the central systems (Fodor, 2000 gives a book-length treatment of this issue).

3.3 Debates about modularity

Fodor's account of modularity has a positive part – input systems are modular – and a negative part – central systems are not modular. The account has been very influential, but also controversial. Many of the criticisms of the positive part of Fodor's account focus on the property of informational encapsulation, which Fodor regards as the essence of modularity.

One important kind of challenge takes the form of presenting evidence that a system that Fodor regards as an input system is not really informationally encapsulated. A defender of modularity might respond by saying that the processing that is influenced by background knowledge and beliefs takes place in the central systems, after the encapsulated input system has delivered its output; the locus of top-down influence is post-perceptual. This response would fit well with the sixth property of modules, namely that their final outputs are *shallow*. But here we must be careful because Fodor actually introduced a further property of input systems, one that we have not previously mentioned. Fodor regards the input systems as delivering information to the central systems so that it can be used in *conscious* inferences. So the response would also need to take into account the idea that the outputs of modules are accessible to consciousness or 'phenomenologically salient' (Fodor, 1983, p.87).

Suppose, for example, that it seemed as though visual object recognition was not informationally encapsulated. A response would be to say that visual object recognition takes place in the central systems and not in the modular visual-input system. In Marr's theory of visual processing (Chapter 3), the primal sketch and the 2½D sketch are shallow levels of representation and, to that extent, they would appear to be suitable candidates for the output of a visual-input system. But for Fodor they do not meet the requirement of phenomenological salience – 'from the point of view of phenomenological accessibility, perception is above all the recognition of objects and events' (Fodor, 1983, p.94) – and so the response would fail.

According to Fodor's account of the mind, the processing in input systems is informationally encapsulated *and* the outputs of input systems are accessible to consciousness (phenomenologically salient). These two requirements may sometimes impose conflicting pressures on the location of the interfaces between input systems and central systems.

3.3.1 Arguments against informational encapsulation

According to Fodor, in the processing of spoken language the detection of irony and the interpretation of metaphor are unlikely to be informationally encapsulated. Those processes belong in the central systems because they require access to our beliefs, including beliefs about the speaker's intentions, about whether what the speaker said would make sense if interpreted literally, and much else. So, at what point, earlier than those processes of pragmatic interpretation, is the interface between the language-input system and the central systems located? Fodor's proposal is that the language-input system delivers information about the linguistic form of the sentence that was uttered – which words it contains and how they are put together. This output is relatively shallow – it certainly falls well short of complete understanding of the message that the speaker was intending to get across – but it is still phenomenologically salient.

In a critique of Fodor's account of modularity, Marslen-Wilson and Tyler (1987) review a range of evidence that seems to show that the language-input system, as conceived by Fodor, is not informationally encapsulated. Processing in the language-input system is, they argue, influenced by knowledge about the non-linguistic world. In one experiment (Marslen-Wilson *et al.*, 1988), participants listened to sentences, some of which were anomalous in various ways. For our purposes, the crucial comparison is between sentences in normal prose, such as 'The boy held the guitar', and sentences containing a pragmatic anomaly, such as 'The boy buried the guitar'. The participants' task was simply to press a button as soon as they detected a previously specified target word – in the case of these examples, the word GUITAR. The results show that participants responded significantly faster when the target word was presented in the normal sentence than when it was presented in the pragmatically anomalous sentence.

This finding is important because the anomalousness of the second sentence does not result from the syntactic properties of BURY and GUITAR, and it is not explained by the literal meanings of those words, as specified within the language-input system. As the authors explain: 'The pragmatic oddness of burying a guitar is something that we have to infer, *given our knowledge of the world*, and given what we know about guitars, the likely effects of burying them, and so on' (Marslen-Wilson *et al.*, 1988, p.6; emphasis added). Thus, the finding provides evidence that knowledge of the non-linguistic world influences on-line processing of spoken language input.

In response to this finding, a defender of modularity might propose that the locus of top-down influence is post-perceptual. In a little more detail, the proposal would have two components. First, it would be argued, spoken word recognition is an informationally encapsulated process that takes place within the language-input module. But second, so the argument would continue, deciding to press the button is not an informationally encapsulated process, and is therefore a matter for the central systems. Even under speeded conditions, the decision depends on the participant's beliefs. A slower response to the word GUITAR when it is presented in the pragmatically anomalous sentence might therefore reflect 'the hearer's inability to believe that the speaker could have said what it sounded like he said' (Fodor, 1985, p.5).

Other experiments reviewed by Marslen-Wilson and Tyler (1987) provide evidence that non-linguistic knowledge influences the resolution of ambiguity in phrases such as 'visiting relatives' or 'landing planes' (Tyler and Marslen-Wilson, 1977) and influences the interpretation of anaphors[2] (Tyler and Marslen-Wilson, 1982; Marslen-Wilson *et al.*, 1993). In each case where informational encapsulation is challenged, the defender of modularity may respond by claiming either (i) that the top-down influence is coming from within the input system and not from the central systems, or else (ii) that the processing that is being influenced occurs in the central systems and not within the input system. And, of course, in each case there can be further debate about whether the response is adequate.

Let us now return to Swinney's (1979) experiment, and the question as to whether *lexical access* is informationally encapsulated. Swinney found that both ANT and SPY are primed when they are presented immediately following BUGS and, as we have seen, these semantic priming effects could be explained in terms of the organization of the mental lexicon, and not in terms of information from the central systems. However, Swinney also discovered that if BUGS is heard in a sentence about insects and ANT or SPY is presented three syllables (about one second) after BUGS, then ANT is primed but SPY is not. This indicates that, while both meanings of BUGS (both words spelled BUGS) are accessed initially, there is then a disambiguation process in which one of them is selected at the expense of the other. It is plausible that this process of selection draws on information from the central systems, but the locus of top-down influence can be argued to occur *after* lexical access. Indeed, Swinney (1979, p.657) suggests that 'a very rapid postaccess decision process is at work'.

---

[2] Anaphors (which may be anaphoric expressions such as pronouns or may be anaphoric devices that are not phonologically explicit in spoken language) refer back to someone or something that has been introduced or identified by an antecedent in the previous text or linguistic context.

Marslen-Wilson and Tyler (1987) also raise some more general issues for Fodor's account of modularity. Fodor connects fast processing in input systems with the properties of being mandatory and being informationally encapsulated. But the influence of non-linguistic knowledge does not make processing slow. In the experiment that we described earlier (Marslen-Wilson *et al.*, 1988), the difference in mean reaction times to target words between the normal and pragmatically anomalous sentences was just 28 msec. So processing that involves central systems can still be fast and speed of processing does not allow us to draw a clear boundary around the modular language-input system. Perhaps surprisingly, it is also not straightforward to use the notion of informational encapsulation – the essence of modularity – to draw the boundary between an input system and the central systems. The reason is that the property of informational encapsulation is defined in terms of the absence of top-down influence from information that is stored *in the central systems*.

Fodor says that each module has a 'proprietary database' of information that the module's computations can draw on (Fodor, 1985, p.3). Informational encapsulation requires only that this data should not include all the information that is stored in the central systems. Domain specificity suggests, in addition, that the proprietary database of the language module, for example, should include only information about language – information about phonology and syntax, and information about words (a lexicon), but not information about whether it is likely that someone would bury a guitar. Against this suggestion, Marslen-Wilson and Tyler (1987) urge that we should not so readily assume that *language processing* follows the contours of *linguistic theory*. More generally, there is still work to be done before we can confidently, and non-arbitrarily, draw boundaries between input systems and the central systems.

3.3.2 Prospects for the psychology of the central systems

The negative part of Fodor's account of modularity is that central systems are not modular and, as we have seen, Fodor is pessimistic about the prospects for a computational psychology of the central systems. Critics of this negative part of the account argue that the mind is modular throughout – *massively modular* (Sperber, 1994, 1996, 2002) – and that, consequently, central cognitive processes can also be brought within the scope of the CMM.

The massive modularity hypothesis draws support from evolutionary psychology (Barkow *et al.*, 1992; Cosmides and Tooby, 1994). The idea is not that there is a module for thinking, another for belief formation, and another for decision making, but rather that there is a host of more specialized modules, each dedicated to the solution of a particular, and perhaps quite

✱ ACTIVITY 20.5

The Marslen-Wilson and Tyler data concern the language-input module. Can you find any evidence – from this book or elsewhere – that is relevant to Fodor's claim that the *visual*-input system is informationally encapsulated?

COMMENT

You might start by looking again at Chapter 3, Section 5 on 'Constructivist approaches to perception'. That section describes evidence supporting the hypothesis that knowledge affects what we perceive. How might Fodor respond?

idiosyncratic, type of problem. Dan Sperber argues that the massive modularity hypothesis:

> *gets support from a wealth of recent work . . . tending to show that many basic conceptual thought processes found in every culture and in every fully developed human are governed by domain-specific competences. For instance it is argued that people's ordinary understanding of an inert solid object, of the appearance of an organism, or of the actions of a person are based on three distinct mechanisms: a naive physics, a naive biology, and a naive psychology.*
>
> (Sperber, 1996, p.123)

In a similar vein, Steven Pinker says, at the beginning of his book, *How the Mind Works*: 'the mind is a system of organs of computation designed by natural selection to solve the problems faced by our evolutionary ancestors in their foraging way of life' (Pinker, 1997, p.x). You may not be surprised to learn that Fodor argues against massive modularity in his book, *The Mind Doesn't Work That Way* (2000; for further discussion, see Pinker, 2005a; Fodor, 2005; Pinker, 2005b).

If massive modularity is to provide grounds for optimism about the prospects for scientific understanding of central cognitive processes then one thing that needs to be demonstrated is how processes such as belief formation, decision making, and analogical reasoning could be carried out by a massively modular cognitive architecture. Ideally, it would be shown how those processes could, after all, fall within the scope of the CMM. A sophisticated defence of massive modularity that does attempt to show these things is provided by Peter Carruthers (2006).

As you consider this debate about the negative part of Fodor's account of modularity, you should be aware that complications could arise from the fact that advocates of massive modularity may not mean the same by the term 'module' as Fodor means. A naïve physics, or biology, or psychology, as described by Sperber, sounds like a body of knowledge. But Fodor's modules are processing systems, rather than stored bodies of domain-specific knowledge. Carruthers relaxes Fodor's account of modules and, in particular, does not require that modules are informationally encapsulated – the essence of Fodorian modularity. So it seems that an advocate of massive modularity might actually accept the negative part of Fodor's account – the central systems are not modular in Fodor's sense – while maintaining that central cognitive processes are underpinned by systems that are modular in a different way.

You would then need to determine where, if anywhere, the real disagreement between the two sides is located (Robbins, 2009). Do the advocates of massive modularity accept that cognitive processes of belief formation have the properties of being isotropic and Quinean? Do they accept that those properties require that the processes are unencapsulated? Do they accept that processes that are unencapsulated are not computationally tractable?

SUMMARY OF SECTION 3

- Fodor proposes that the mind can be divided into sensory transducers, modular input systems, and non-modular central systems.
- Modular input systems are characterized by a cluster of properties, but the properties of being domain specific and informationally encapsulated are especially important.
- Informational encapsulation means that the processing undertaken by a module is not affected by knowledge stored in the central systems. Top-down information flow within a module is permitted.
- The central systems receive output from the modular input systems and are involved in belief formation and decision making. They are domain general and unencapsulated.
- Fodor proposes that a good model for the functioning of the central systems is the kind of non-demonstrative inference that is used in scientific confirmation – that is, abductive inference

or inference to the best explanation. This kind of inference has the properties of being isotropic and Quinean.

- Fodor's 'First Law of the Non-existence of Cognitive Science' warns that the prospects for a computational psychology of the central systems are bleak.
- Fodor's account of modularity has a positive part – input systems are modular – and a negative part – central systems are not modular.
  - Critics of the positive part of Fodor's account of modularity have questioned whether input systems are really informationally encapsulated.
  - Critics of the negative part of Fodor's account have proposed that the mind is modular throughout.

4 COGNITIVE NEUROPSYCHOLOGY

We now turn to a branch of cognitive psychology that relies on an assumption of modularity to draw conclusions about cognitive processes. Cognitive neuropsychology uses data from patients with acquired cognitive disorders following brain injury in order to constrain models and theories of normal cognition (Chapter 13). The term 'neuropsychology' is also used for a discipline that is concerned with the relationship between brain and behaviour (and is thus close to behavioural neurology). Cognitive neuropsychology and neuropsychology in this other sense have common antecedents in nineteenth-century neurology. But cognitive neuropsychology is different from neuropsychology in being an approach or research programme within cognitive psychology. Its subject matter is not the brain, but the information-processing structure of the mind.

For cognitive neuropsychology, the most important pattern of impaired cognitive functioning in patients is **double dissociation** of impairments on two cognitive tasks: one patient (A) shows impaired performance on Task 1 but performs at a normal or near normal level on Task 2, while a second patient (B) shows the reverse pattern, impaired on Task 2 but unimpaired or near normal on Task 1. From this evidence, it is inferred that the two tasks involve separate information-processing routes or mechanisms.

In order to draw these inferences about the structure of normal cognition, cognitive neuropsychology relies on several assumptions:

1. The first of these is an assumption of *modularity*: the normal cognitive system is made up of somewhat independent components or information-processing mechanisms. However, the notion of modularity that is used in the practice of cognitive neuropsychology has fewer commitments than Fodorian modularity. Two points of contact between the modularity assumption in cognitive neuropsychology and issues in Section 3 of this chapter are briefly described in Box 20.2 and Box 20.3.
2. The second assumption is *universality*: the modular structure or *functional architecture* of the mind is the same, in relevant respects, for all normal (neurologically healthy) individuals (Caramazza, 1986).
3. The third assumption is that, when one cognitive component is damaged, this does not bring about substantial reorganization of the prior modular structure. The undamaged components continue to operate as before and no new components are added. This assumption is sometimes called *transparency* (Caramazza, 1986) and sometimes *subtractivity* (Coltheart, 2001). (See Chapter 13 for further discussion.)

BOX 20.2 Modularity in central systems

Fodor's main argument for the negative part of his account of modularity was reviewed in Section 3.2. The central systems are not modular because central cognitive processes, such as belief formation, have properties – being isotropic and Quinean – that require them not to be informationally encapsulated. But, if the notion of modularity that is used in cognitive neuropsychology is closer to the everyday notion than to Fodor's, then there may be modules – in the sense of somewhat independent components – in the central systems. Damage to separate components of the central systems might, in principle, result in selective cognitive impairments and, indeed, there is evidence that memory systems can be selectively impaired, that the ability to do arithmetical calculation can be destroyed, that attentional systems can be damaged, and that everyday psychological understanding (theory of mind) can be impaired (Shallice, 1988; Ellis and Young, 1998; and Baron-Cohen *et al.*, 1999 describe evidence of all these impairments).

One of the properties associated with Fodorian modularity is fixed neural architecture and Fodor suggests that this property, like informational encapsulation, is missing from the central systems and that, perhaps, 'central problem-solving is subserved by equipotential neural mechanisms' (Fodor, 1983, p.119). Here, Fodor alludes to Karl Lashley (1930), who maintained that some functions (such as a rat's capacity to run through a maze) are localized in the whole cortex and can be impaired by damage anywhere in the cortex. If several separate components of the central systems were each to be localized in the whole cortex then it would be difficult for them to be separately damaged and selective impairments would not be expected. On the face of it, Fodor's suggestion that the neural substrate of central cognitive functions is diffuse may be in tension with the evidence of selective central cognitive impairments. (One possibility to consider would be that the neural basis of central cognitive *processing* is diffuse but that domain-specific *knowledge modules* can be separately damaged.)

There is also a more general issue here. Cognitive neuropsychology is a division of cognitive psychology and its subject matter is the information-processing structure of the mind. But the practice of cognitive neuropsychology depends on an additional assumption about the brain – an assumption of (at least some) *anatomical modularity*. If selective impairments are to be explained in terms of damage to components of the mind – components of input systems or of the central systems – then components that are functionally separate (modules) must be able to be damaged separately. This would not be possible if many separate modules all shared the same neural substrate.

4.1 Inference in cognitive neuropsychology

In cognitive neuropsychology, inferences about the modular structure of the normal cognitive system are drawn from patterns of cognitive impairment in patients, and the most important pattern is double dissociation. How do those inferences work and why is double dissociation regarded as evidence for separate cognitive modules? We can answer this question in three steps.

First, *why is evidence from brain-damaged people relevant to models of normal functioning at all?* There is no *a priori* way of putting a boundary around the kinds of evidence that may be relevant to the confirmation or disconfirmation of a hypothesis (see our earlier discussion of hypothesis confirmation as being isotropic). So there is no *a priori* reason to suppose that evidence from brain-damaged people will or will not be relevant to models of normal cognitive functioning. This is an empirical issue. However, if the *modularity* assumption about the normal cognitive system is correct then there may be separate

BOX 20.3 A non-Fodorian definition of modularity

Coltheart (1999) proposes that a module should be defined as 'a cognitive system whose application is domain specific' (1999, p.118) and that it should be left as an empirical question whether a particular domain-specific module has any of the other properties of Fodorian modules, such as having fixed neural architecture or being informationally encapsulated.

The notion of a module as a domain-specific processing system may well be attractive to cognitive psychologists who believe that input systems, such as the language-input system, are not informationally encapsulated. Also, Coltheart's notion has an advantage of generality, in that it may apply to some central cognitive systems as well as to input systems. But, if the notion is to be put to use, then it requires a good account of what domain specificity is – and what a domain is. Fodor himself warns that we have to be careful to avoid a 'trivial kind of domain specificity' (Fodor, 1983, p.48). Someone might say that cows and sheep are two different domains and that cow perception and sheep perception are domain-specific processes. But this would be a purely verbal manoeuvre if the same mechanisms of visual perception are involved in perceiving both cows and sheep.

Coltheart recognizes the need to provide an account of domain specificity that is empirically based, rather than merely verbal. So he invites us to imagine that we started out with the idea that there is a single visual-recognition system for faces, objects, and written words: 'Then we noticed that in the neuropsychological literature there were reports of patients with impaired visual word recognition but who retained face recognition and of patients with impaired visual word recognition but who retained visual object recognition. We also noticed reports of patients with impaired visual object recognition but who retained face recognition or visual word recognition . . . This collection of results refutes our original idea . . . Instead, it suggests that there are three separate modules' (Coltheart, 1999, p.119).

This is a typical example of inference in cognitive neuropsychology. Case studies of patients reveal dissociation of impairments between tasks – in this case, multiple dissociations amongst the three tasks of recognizing faces, objects, and written words. This evidence supports the hypothesis that the three tasks involve three separate information-processing systems, each of which is domain specific. Faces, objects, and written words – unlike cows and sheep – are empirically (and not just verbally) different domains. Thus, Coltheart's notion of a module as a domain-specific processing system seems to fit well with the practice of cognitive neuropsychology.

cognitive modules involved in different cognitive tasks. If the assumption of *anatomical modularity* is also correct, then the separate cognitive modules may be independently damaged. If the subtractivity/transparency assumption is correct, then damage to one module will result in impairment of functions that involve that module, while other functions are unimpaired (or spared). So the cognitive neuropsychology framework has the potential to explain patterns of selective impairment in patients following brain injury. If a hypothesis about separate modules in the normal cognitive system allows the *best explanation* of a pattern of impairment then the hypothesis is, to that extent, supported – by inference to the best explanation.

It might be, of course, that these assumptions are incorrect. Brain injury might, in principle, have resulted in apparently random patterns of impairment that could not be understood at all in terms of models of normal cognition. But it is a consistent neuropsychological finding, from the nineteenth century until the present, that brain injury often results in selective impairment of specific cognitive functions.

Second, *why does double dissociation provide better evidence for separate modules than a single (one-way) dissociation?* If the *universality* assumption is correct, then the modular structure of the normal cognitive system is the same in all neurologically healthy individuals. Suppose that, in healthy individuals, two separate modules, M1 and M2, are involved in two different tasks, Task 1 and Task 2. Then independent damage to each of the modules in two patients following brain injury provides a good explanation of double dissociation of impairments on the two tasks. This good explanation may turn out to be the best explanation.

Suppose that there were only a one-way dissociation: a patient shows impaired performance on Task 1 but performs at a normal level on Task 2. One possible explanation would be that two separate modules, M1 and M2, are involved in the two tasks and that, in this patient, M1 is damaged while M2 is intact. But there is also a possible alternative explanation. It might be that just the same modules are involved in the two tasks but that Task 1 is more difficult – more demanding of cognitive resources – than Task 2. If that were so then, after partial damage to the modules that are involved in both tasks, performance on Task 1 might be impaired while performance on Task 2 remained normal.

A double dissociation, in which a second patient shows the reverse pattern (normal performance on Task 1 but impaired performance on Task 2), is better evidence for separate modules than the single dissociation because it excludes the possible alternative explanation in terms of the greater difficulty of one of the tasks. But, of course, excluding one possible alternative explanation is very far from providing a logical guarantee that the hypothesis of separate modules involved in the two tasks is correct. Normal science proceeds by inference to the best explanation and evidence does not provide logical guarantees but, at best, a change in the balance of probabilities between competing theories or models (Coltheart and Davies, 2003; Davies, 2010).

Third, just as impairments can be dissociated, they can also be associated. *If impairments on two tasks usually occur together, can we infer that just the same modules are involved in the two tasks?* The hypothesis that, in the normal cognitive system, the same modules are involved in Task 1 and Task 2 certainly provides one possible explanation of the co-occurrence of impairments on the two tasks following brain injury. But there is a fairly obvious possible alternative explanation in terms of neuroanatomy. Even if functionally separate modules, M1 and M2, were involved in the two tasks, it might be that brain injury that damaged one of these modules would inevitably damage the other one as well. The two modules might be localized in the same small region of the brain, or their positions relative to a blood vessel in the brain might make it overwhelmingly likely that a blockage in or bleeding from that blood vessel would damage both modules. In short, there might not be sufficient anatomical modularity to allow the two modules to be damaged independently.

4.2 Cognitive neuropsychology and the past-tense debate

In this section we return to the past-tense debate (Section 2.3) and consider how neuropsychological evidence has contributed to the debate.

4.2.1 Double dissociation evidence

Evidence of double dissociation on tasks related to the past tense of regular and irregular verbs would support the hypothesis that separate modules are involved in the formation of the past tense of regular and irregular verbs. It would favour dual-route models, such as the words and rules model, over single-route connectionist models (e.g. Rumelhart and McClelland, 1986; Plunkett and Marchman, 1993).

Double dissociations were described in two papers published in 1997. Ullman and colleagues (1997) used an experimental paradigm in which participants read aloud pairs of sentences, such as, 'Every day I *dig* a hole. Just like every day, yesterday I ________ a hole', and filled in the blank in the second sentence. Patients suffering from Alzheimer's disease with severely impaired memory for words performed the task significantly worse (by comparison with control participants) for irregular verbs (e.g. DIG) than for regular verbs (e.g. LOOK) or novel verbs (e.g. PLAG). Patients with Parkinson's disease and severely suppressed movement (hypokinesia) showed the reverse pattern, performing significantly worse (by comparison with control participants) for regular verbs, and much worse for novel verbs, than for irregular verbs. Ullman and colleagues argue that the double dissociation supports 'psycholinguistic theories that emphasize grammar and lexicon as distinct components … especially in the treatment of regular and irregular grammatical phenomena' (1997, p.274; see also Ullman *et al.*, 2005).

Marslen-Wilson and Tyler (1997, 1998) used an auditory priming paradigm that did not require participants to read text or give spoken responses. Participants listened to a spoken target word, immediately preceded by a prime. The prime was either morphologically related to the target (JUMPED–JUMP; GAVE–GIVE), semantically related to the target (GOOSE–SWAN), or unrelated to the target. Filler stimuli included non-words, and the participants' task was simply to make a lexical-decision response ('Is this a word or not?'). The crucial comparison was between priming of regular verbs by their past-tense forms (e.g. JUMPED–JUMP) and priming of irregular verbs by their past-tense forms (e.g. GAVE–GIVE). In this primed lexical decision paradigm, neurologically healthy individuals responded faster to targets when the prime was morphologically or semantically related than when the prime was unrelated, and both regular and irregular verbs were primed by their past-tense forms.

However, three patients did show significant differences in priming between regular verbs and irregular verbs. Two patients, with impaired comprehension and production of inflected forms following left-hemisphere stroke, exhibited priming for irregular verbs but not for regular verbs. A third patient, with symptoms of Broca's aphasia following damage to both hemispheres, exhibited the reverse pattern: priming for regular verbs but not for irregular. Marslen-Wilson and Tyler argue that this double dissociation suggests 'the two morphological categories ally themselves with different types of mental computation' (1997, p.593).

4.2.2 The connectionist response

This evidence of double dissociation supports dual-route, rather than single-route, models of formation of the past tense. But it certainly does not provide a logical guarantee that the words and rules model is correct. It does not even provide a logical guarantee that some dual-route model is correct and all single-route models are incorrect.

Kim Plunkett and colleagues (Juola and Punkett, 2000; Plunkett and Bandelow, 2006) have made this point vivid by demonstrating that double dissociation of impairments may arise from *random* damage to a *single-route* network. They trained a single-route network to form the past tense of verbs (and the plural of nouns). The network (analogous to a neurologically healthy individual) was then copied more than 13,000 times, and each copy subjected to random damage (by removal of 1 per cent of its connections). Amongst these 'lesioned' networks (analogous to patients following brain injury) there were pairs of networks instantiating double dissociation of impairments between regular and irregular verbs. In short, double dissociation evidence from damaged networks may be compatible with the undamaged network having only a single route.

Joanisse and Seidenberg (1999) constructed a single-route connectionist model of past-tense formation that was different from earlier connectionist models in an important way: it included semantic representations of verbs as well as phonological representations. By subjecting components of the model to damage, Joanisse and Seidenberg were able to simulate the double dissociation that Ullman and colleagues (1997) found in patients suffering from Alzheimer's disease and Parkinson's disease. When the semantic representations in the trained network were damaged, the effect was greatest for irregular verbs (as in the patients with Alzheimer's disease); when the phonological representations were damaged, the effect was greatest for novel verbs (as in the patients with Parkinson's disease). Thus, in opposition to the words and rules model, Joanisse and Seidenberg defend a single-route theory of past-tense formation: 'The same network structure is used in processing all words. …

The explanation for the behavioral dissociations lies with the fact that the network includes distinct phonological and semantic representations' (1999, p.7597; see also Joanisse and Seidenberg, 2005 for a similar account of neuroimaging evidence).

4.2.3 Cognitive neuropsychology and connectionist modelling

Twenty-five years or more after Rumelhart and McClelland published their connectionist model of the English past tense, the debate continues. It may be that the empirical evidence will not adjudicate decisively between the competing accounts of past-tense formation and that dual-route words and rules models and single-route connectionist models will do about equally well in accounting for the data – not only data from behavioural experiments but also neuropsychological and neuroimaging evidence. (For a recent connectionist model and new behavioural evidence, see Woollams *et al.*, 2009.)

There is a parallel debate about reading words aloud. Again, there are competing models – the dual-route DRC model (Coltheart *et al.*, 2001) and the single-route connectionist models (Seidenberg and McClelland, 1989; Plaut *et al.* 1996). Here, too, double dissociation – between surface dyslexia and phonological dyslexia – has played an important role. Surface dyslexia (impaired reading of irregular words while reading of regular words and non-words is relatively spared) is explained in terms of damage to the lexical route in the DRC model, and phonological dyslexia (impaired reading of non-words while reading of regular and irregular words is relatively spared) is explained in terms of damage to the non-lexical route, with its spelling-sound rules. However, a single-route connectionist model (Plaut *et al.*, 1996) explains the double dissociation in terms of damage to semantic representations or to phonological representations.

In both these areas – the past tense and reading aloud – continuing work is providing new evidence against which competing models can be assessed. Aside from asking which kind of model, rule-based or connectionist, can better fit the data, researchers in these areas are asking another question. Which kind of model offers the *better explanation* of the available evidence – which style of explanation has broader scope and offers better prospects for unifying seemingly disparate phenomena, which is more parsimonious, which achieves greater depth, which leads to the more interesting new predictions? These issues about fitting and explaining data are at the forefront of contemporary debate (Coltheart, 2006a; Seidenberg and Plaut, 2006).

SUMMARY OF SECTION 4

- Cognitive neuropsychology uses data from patients with acquired cognitive disorders following brain injury in order to constrain models and theories of normal cognition.
- For cognitive neuropsychology, the most important pattern of impaired cognitive functioning in patients is double dissociation of impairments on two cognitive tasks.
- Neuropsychological evidence of double dissociation of impairments has contributed to the past-tense debate because it appears to support dual-route models over single-route models.
- Connectionist modellers have demonstrated that double dissociation of impairments may result from random damage to a single-route model.
- Double dissociation of impairments may also result from damage to separate semantic and phonological representations in a connectionist network, even though regular and irregular verbs are processed through the same connections.
- In the past-tense debate, and in a parallel debate about reading aloud, researchers ask not only which kind of model can fit the data better, but also which offers the better explanation of the available evidence.

5 COGNITIVE PSYCHOLOGY AND THE BRAIN

In this final section, we turn to one aspect of the relationship between mind and brain – a large topic with a very long history reaching back at least to ancient Greek philosophy, more than 2,000 years ago. The specific aspect that we shall discuss is the relationship between cognitive psychology and the study of the brain – that is, neuroanatomy or the study of neural circuits (we shall call this neurobiology in what follows).

5.1 Levels of explanation

If you look back over the previous chapters in this book, you will find relatively little discussion of the way in which cognitive systems are actually implemented by the brain. Why is this? There are, we suggest, two related reasons. The first is the influence of David Marr's (1982) account of the three levels at which an information-processing system can be understood. As you saw in Chapter 1, these three levels are:

- **Level 1**: The level of computational theory, where we ask what is being computed, and why it is being computed. When we know what is being computed, we can also assess how, in principle, it could be computed.
- **Level 2**: The level of representation and algorithm, where we ask how the computation that is abstractly described at level 1 is actually carried out in the system that is under investigation. More specifically, we ask in what formats input information and output information are represented, and by what algorithm the transformation from input to output is accomplished.
- **Level 3**: The level of hardware implementation, where we ask how the representations and algorithms specified at level 2 are physically realized. When we are trying to understand information processing in human cognitive systems, we ask how the representations and algorithms specified at level 2 are physically realized in the human brain.

Experimental research in cognitive psychology is generally aimed at understanding actual representational structures and cognitive processes and so it is most directly concerned with Marr's second level. But any theory at the second level is *constrained from above* by the computational theory at the first level. An empirical proposal about how a computation is actually carried out will be incorrect if the proposed algorithm is not even a possible way, in principle, of accomplishing the transformation from input to output. Theories at the second level are also *constrained from below* by the theory of hardware implementation, or physical realization, at the third level. As Marr put it: 'Some types of algorithm will suit some physical substrates better than others' (Marr, 1982, p.24).

Marr's overall view, which has been influential in cognitive science and cognitive psychology, was that constraint from above, by the computational theory, took precedence over constraint from below, by the hardware implementation. Reflecting on neurophysiological investigations of vision in the 1970s, he became convinced that an important level of understanding had been missing:

> *There must exist an additional level of understanding at which the character of the information-processing tasks carried out during perception are analyzed and understood in a way that is independent of the particular mechanisms and structures that implement them in our heads. This was what was missing – the analysis of the problem as an information-processing task. Such analysis does not usurp an understanding at other levels – of neurons or of computer programs – but it is a necessary complement to them, since without it there can be no real understanding of the function of all those neurons.*
>
> (Marr, 1982, p.19)

The second reason why there has been relatively little discussion of the organization and function of the physical brain in cognitive psychology is that, until recently, there have been few techniques and tools that could be used to undertake the kinds of investigation at level 3 that would speak to issues concerning human cognitive functioning. In essence, experimental investigations were limited to animal studies, where invasive methods such as lesion studies and single-cell recording of neurons could take place. For ethical reasons, these kinds of studies could not be carried out on humans.

However, the situation has changed quite dramatically over recent years. Neuroimaging techniques such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and magnetoencephalography (MEG) allow metabolic, electrical, and magnetic correlates of neural activity to be measured, with varying degrees of spatial and temporal resolution, during the performance of cognitive tasks (Chapter 14). At the same time, there has been massive growth in connectionist modelling that is claimed to be 'brain-like' or at least neurally plausible. So the question of the relationship between the kinds of cognitive models discussed in this book and neural models that are pitched at Marr's level 3 has become a matter of significant theoretical interest and, indeed, dispute.

Contemporary theorists in the cognitive sciences commonly and reasonably assume that cognitive functioning is realized in neural functioning. So Marr is surely right that, in principle, things we find out about neural functioning might constrain models and theories of cognitive processes. This bottom-up constraint from level 3 to level 2 goes hand-in-hand with top-down guidance from level 2 to level 3. It is hard to see how an investigation at the level of the neural implementation of cognitive processes could ever get started without there first being a model (however sketchy) of the cognitive processes themselves.

If there is both bottom-up constraint and top-down guidance between Marr's second and third levels then we should expect reciprocal interaction between cognitive psychology and neurobiology, with challenges and insights flowing in both directions. This, in outline, is what Patricia Churchland (1986) describes as **co-evolution of theories**: 'theories at distinct levels often co-evolve . . . as each informs and corrects the other' (P.S. Churchland, 1986, p.284).

We should not, however, expect that drawing inferences about cognitive psychological theories from neurobiological findings will be easy or straightforward. A simple example illustrating this point is presented in Box 20.4.

5.2 Co-evolution and reduction

Patricia Churchland links co-evolution of theories with reduction of one theory to the other: 'the discoveries and problems of each theory may suggest modifications, developments, and experiments for the other, and thus the two evolve towards a reductive consummation' (P.S. Churchland, 1986, p.284). Even if it is not really obligatory that interdisciplinary interaction should aim at 'reductive consummation', the idea that theories in cognitive psychology *might* reduce to theories in neurobiology requires some discussion.

The topic of inter-theoretic reduction is a complex one (Churchland and Churchland, 1998 give a basic introduction). But the basic thought is that 'because everything in the world is made of the same basic stuff in complex combinations, the laws of biology ought to be derivable from those of chemistry, and the laws of chemistry from the laws of physics' (Ladyman, 2002, p.95). Similarly, the idea is that the laws of cognitive psychology ought to be logically derivable from laws that figure in neurobiological theories. Examples from the history of science that are usually adduced to illustrate and motivate this view are the reduction of genetics to molecular biology and the reduction of chemistry to quantum mechanics.

5.2.1 Reduction and the CMM

There is, however, a line of thought leading to the opposite claim, that cognitive psychology is not even a candidate for reduction to neurobiology. The CMM

BOX 20.4 Descending pathways and informational encapsulation

As we saw earlier in this chapter, Jerry Fodor argues that perception is informationally encapsulated from our beliefs and knowledge about the world. In one of a number of responses to this claim, the neurophilosopher Paul Churchland pointed out that, in addition to neuronal pathways that 'ascend' from the retina to primary visual cortex (via structures such as the lateral geniculate nucleus), there is evidence that there are also matched 'descending pathways' that 'lead us stepwise back through the intermediate brain areas and all the way out to the earliest processing systems at the retina' (P.M. Churchland, 1989, p.266). Churchland interprets these 'descending pathways' as feedback pathways from the central systems to the input systems that 'strongly suggest' that perceptual processes are *not* encapsulated.

Fodor's reaction to this neurobiological objection to informational encapsulation is instructive. He says: 'Heaven knows what psychological function "descending pathways" subserve. . . . One thing is clear: if there is no cognitive penetration of perception then at least "descending pathways" aren't for that' (Fodor, 1990, p.261). One way to interpret this remark is to read Fodor as denying, in principle, the relevance of neurobiological findings to cognitive psychology. But this would not be a credible interpretation, given Fodor's repeated insistence that there is no *a priori* way to put a boundary around the kinds of evidence that may be relevant to a theory. (Recall the isotropy of scientific confirmation discussed in Section 3.2.)

What Fodor is suggesting – in accord with the co-evolution view – is that *as matters currently stand* our understanding of the role of the descending pathways is too primitive to be taken as evidence against informational encapsulation. Moreover, we should expect that empirical theories in cognitive psychology (and, indeed, abstract computational theories) will guide our understanding of the role of those descending pathways. If we were to find good reasons for thinking that those pathways do carry information from central systems to input systems, then we might need to rethink our understanding of modules and encapsulation. But until then, assumptions of modularity and encapsulation may guide our understanding of the functions of descending pathways.

models 'the mind as the software of the brain' (Block, 1995). The same software, with the same scheme of representations and the same algorithmic processes (Marr's level 2), could run on different hardware (Marr's level 3). So the same cognitive psychological theory could be true of minds with very different physical realizations, some of them biological but some of them not. This point about the **variable physical realization** of cognitive states and processes blocks any straightforward reduction of cognitive psychology to neurobiology. A truly general cognitive psychological theory would apply to the cognitive states and processes of beings with wholly different physical compositions, some of which would not be biological at all. No neurobiological theory could match the generality of a psychological theory of that kind. Thus, there is a sense in which, as Block puts it, 'the computer model of the mind is profoundly *unbiological*' (1995, p.390).

This is an important line of thought but we should also notice its limitations. Variable physical realization blocks the reduction of completely general psychological theories, such as the CMM, to neurobiology. But this line of thought, by itself, leaves it open that a psychological theory of specifically *human* information processing might, in principle, be reduced to human neurobiology.

5.2.2 Reduction and the history of science

The linguist Noam Chomsky (2000, 2002) has articulated another problem for reductionism. He argues that inter-theoretic reduction is historically rare, and that where there have been genuine cases of reduction – such as that of chemistry to physics – this was only possible when there was a radical change in the more basic science – in this case, physics. The moral Chomsky draws is that reduction is neither to be aimed for nor to be expected.

The only sensible aim is for each theoretical enterprise to pursue its own path (as did chemistry and physics). It cannot be ruled out that there will be what Chomsky calls the 'unification' of theories (as with chemistry and physics), but nor can we be at all certain that human intelligence will be up to the task of unification. So far as psychology and neurobiology are concerned, 'one can entertain the idea that "the mental is the neurophysiological at a higher level", but for the present, only as a guide to enquiry, without much confidence about what "the neurophysiological" will prove to be' (Chomsky, 2003, p.265).

Chomsky's point can, perhaps, be put in the following way. The reduction of one theory to another requires that both theories are successful in their respective domains. If either (or both) is not antecedently successful, then why would one either expect or want reduction? With regard to psychology and neurobiology, our theories are tentative and far from being in the state where a reduction is, even in principle, in the offing. So, as things currently stand, the claim that psychology will reduce to neurobiology is merely a *historical speculation*, and one based on very scant evidence from the history of science.

5.3 Is cognitive psychology just a historical staging post?

Despite these reservations about reduction, it may still be tempting to think that contemporary cognitive psychology is a mere historical staging post *en route* to the terminus of a neurobiological theory of the mind. You might expect that cognitive psychology will disappear when that end point is reached. Philosophers Ian Gold and Daniel Stoljar have described this view as a version of *the neuron doctrine*: 'a successful theory of the mind will be a wholly neuroscientific theory' (Gold and Stoljar, 1999, p.809).

The neuron doctrine has different versions depending on how the term 'neuroscientific' is interpreted. If a neuroscientific theory is a theory in *cognitive* neuroscience, drawing on some combination of concepts from cognitive psychology and neurobiology (or, more generally, from the psychological and biological sciences), then the neuron doctrine is somewhat bland and uncontroversial. But if a neuroscientific theory is a theory that draws only on neuro*biology* then we have a much more radical neuron doctrine: 'a successful theory of the mind will be a theory of the brain expressed *in terms of* the basic structural and functional properties of neurons, ensembles or structures' (Gold and Stoljar, 1999, p.814). In a little more detail, the radical neuron doctrine says:

> *neurophysiology, neuroanatomy, and neurochemistry will by themselves eventually have the conceptual resources to understand the mind and, as a consequence, a successful theory of the mind will make no reference to anything like the concepts of . . . the psychological sciences as we currently understand them.*
>
> (Gold and Stoljar, 1999, p.814)

On this view, we shall have not the co-evolution of theories, but rather the gradual extinction of cognitive psychological theories as neurobiological theories take over.

A temptation to believe the radical neuron doctrine may arise from an assumption that (as we noted earlier) most cognitive psychologists and cognitive scientists share; namely, that cognitive functioning is realized in neural functioning. It may seem obvious that, if mental functioning is the result of the physical organization and functioning of the brain, then a theory of the mind will ultimately be produced that talks only about the brain – about neurons, neural circuits, neurotransmitters, and so on.

But this would be to move much too fast. To use an example of Gold and Stoljar's (1999, p.815): earthquakes are made up of the movement of physical particles that behave in accordance with the laws of physics. But the science of earthquakes shows no signs of being replaced by physics, and the reason for this is quite general. Just because phenomena of some kind, Xs, are made up of lots of Ys, it does not follow that we will achieve an understanding of Xs in terms of the science of Ys. We should not, then, allow ourselves to slide from the widely held assumption that the mind is realized in the brain to the radical and controversial version of the neuron doctrine. Furthermore, if we were to begin on this slide then it would be completely unmotivated to maintain that the *real* science of the mind is neurobiology. For neural phenomena are made up of the activity of lots of molecules and atoms, protons, neutrons, and electrons, and ultimately, of whatever turn out to be the fundamental building blocks of all matter.

5.4 Is cognitive psychology autonomous from neurobiology?

Philosophical commentators on the relationship between cognitive psychology and neurobiology have used two different notions of autonomy, each of them important.

One notion of autonomy is *irreducibility*. Suppose, as we assume is true, and this book is predicated on, that theories in cognitive psychology provide genuine empirical explanations. If the laws that figure in those theories are not derivable from laws in neurobiology then cognitive psychology is autonomous from neurobiology in this first sense (Fodor, 1998, p.9). If, instead, the key concepts in cognitive psychology can be defined in terms of concepts in neurobiology, and if the laws that figure in cognitive psychological theories are derivable from laws of neurobiology, then cognitive psychology is reducible and not autonomous in that same sense.

Suppose, for a moment, that cognitive psychology were to be reducible to neurobiology. It is important to notice that it would not immediately follow that the radical neuron doctrine was also true. According to the radical neuron doctrine, cognitive psychology will, in due course, be discarded. But a successful reduction can be a way of vindicating, rather than eliminating, the reduced theory. The Churchlands, who are certainly friends of reduction, say: ‘The discipline of psychology will still be with us a hundred years from now, and five hundred, and a thousand. . . . Its existence is unthreatened’ (P.M. Churchland and P.S. Churchland, 1996, p.220).

A second notion of autonomy is *independence*. Cognitive psychology is autonomous from neurobiology in this second sense if it is entirely unconstrained by neurobiology. Note that autonomy in this second sense entails autonomy in the first: if cognitive psychology were to be fully independent from neurobiology then it could not be reducible to neurobiology. However, the reverse does not hold: cognitive psychology could be constrained by neurobiology – as in Marr’s account – without being reducible to neurobiology.

We can therefore reject the autonomy of cognitive psychology in the second sense without rejecting autonomy in the first sense. We can allow that theories in cognitive psychology and neurobiology co-evolve without being committed to reductive consummation. Indeed, this seems to be quite an attractive account of the relationship between the two disciplines (Stone and Davies, 1999).

Some psychologists maintain that the cognitive psychological level (Marr’s level 2) takes *priority* over the neurobiological level (Marr’s level 3). This priority has theoretical and practical aspects. The theoretical aspect of the priority of cognitive psychology is expressed by Coltheart and Langdon, when they say that ‘it can be very hard to understand what a system is actually doing if one’s only information about it is a description at the physical-instantiation level’ (1998, p.150). This aspect of the priority of cognitive psychology seems to coincide with that we earlier described as top-down guidance from Marr’s level 2 to level 3. In Marr’s account, the project at level 3 is to discover how the representations and algorithms *specified at level 2* are physically realized in the brain. It would seem to be a very much more difficult project to work out which representations and algorithms were being realized, given only a neurobiological description of neural states and processes.

The practical aspect of the priority of cognitive psychology is simply a manifestation of the fact that all scientists need to make decisions about how to conduct research with limited resources. One option for a cognitive psychologist may be to push ahead with the development of an information-processing model while abstracting, for the time being, from debates within neurobiology. For example, Andy Young describes how the development of a cognitive model of face recognition proceeded in relative abstraction from debate about a neuroanatomical proposal (Bauer, 1984) concerning two of the routes in the model. Young comments: 'It is a good example of how, even though they can often be usefully combined, psychological and neurological hypotheses can have some degree of independence from each other' (Young, 1998, p.44). Cognitive psychologists who make these practical decisions would typically acknowledge, of course, that the model they develop in abstraction from neurobiology is still potentially constrained from below. The model would have to be rejected if there were no neurobiological account consistent with it.

Sometimes, however, these claims about the theoretical and practical priority of cognitive psychology seem to be associated with the less plausible claim that neurobiological findings are irrelevant to cognitive models. Examples of findings whose relevance sometimes seems to be contested include findings about the location of neural activity in neurologically healthy individuals who are performing a cognitive task, or the location of lesions in patients who can no longer perform a particular task. As was suggested in Chapter 1, the relevance of neuroimaging to cognitive psychology has thus become a topic of lively discussion and debate (Henson, 2005; Coltheart, 2006b).

SUMMARY OF SECTION 5

- According to Marr's account of levels of explanation, models in cognitive psychology are constrained from above by the computational theory, and from below by the theory of hardware implementation. It was also part of the account that constraint from above takes precedence over constraint from below.
- On the co-evolution of theories view, there is reciprocal interaction between cognitive psychology and neurobiology, with challenges and insights flowing in both directions.
- There are theoretical and historical grounds to doubt whether cognitive psychology will be reduced to neurobiology.
- The radical neuron doctrine predicts that cognitive psychology will eventually be replaced by neurobiology, but this doctrine is unmotivated.
- Claims about the theoretical and practical priority of cognitive psychology are sometimes combined with less plausible claims about the irrelevance of neurobiological findings to cognitive models.

6 CONCLUSION

This chapter has introduced you to some of the main theoretical debates in cognitive psychology and in cognitive science more broadly. The first debate (Section 2) was about rules and representations and the connectionist challenge to the computational model of the mind (CMM). One strand in this large-scale debate concerned the competing connectionist and rule-based accounts of a particular cognitive process, the formation of the past tense of verbs. The second large-scale debate (Section 3) was about the modularity of the mind and about challenges to the two parts – positive and negative – of Fodor's account of modularity. One important question was whether language processing, and specifically the process of lexical access, is subject to top-down influence from non-linguistic knowledge.

In Section 4, we introduced the debate about the role of double dissociation evidence in cognitive neuropsychology. In this context, returning to the past tense and to connectionist models of past-tense formation allowed us to highlight two points. First, cognitive neuropsychology, like all of normal science, proceeds by inference to the best explanation. Second, competing models of cognitive processes should be assessed not only for how well they fit a particular set of data, but also for how well they explain the available evidence. In the final section, we briefly introduced debates about the relationship between cognitive psychology and neurobiology – reduction, co-evolution, priority, and independence. The contribution that neuroimaging can make to cognitive psychology is a topic of continuing and widespread discussion.

None of these debates has a resolution that carries a consensus amongst cognitive psychologists and cognitive scientists. But perhaps there would be widespread agreement that the development of explicit, implemented computational models – including connectionist models – has made a substantial contribution to cognitive psychology over the last 25 years. And perhaps the success of computational cognitive psychology serves to highlight the question of whether, as Fodor claims, some of the aspects of human cognition that are the most important to us are computationally intractable.

There are, of course, many other theoretical debates that warrant attention. Among those that we have most reluctantly omitted are debates about the cognitive basis of our everyday psychological understanding of ourselves and others (theory of mind; Davies and Stone, 1995), about the prospects for a science of consciousness (Chapter 18; Davies and Humphreys, 1993; Weiskrantz and Davies, 2008), and about the extension of the methods of cognitive neuropsychology to disorders, such as delusions, that were previously regarded as psychiatric phenomena (Chapter 15; Stone and Young, 1997).

We hope that you will take away from this chapter a greater insight into the fundamental principles that lie behind theories in contemporary cognitive psychology, and that this will enhance your understanding of the individual topics covered in this book.

Dedication

Tony Stone, the author of the original version of this chapter (published in the first edition of this book in 2005), died in June 2010 at the age of 52. This revised chapter is dedicated to his memory.

FURTHER READING

When we introduced the computational model of the mind, we adopted Ned Block's phrase, 'the mind as the software of the brain'. Block's chapter of that title provides an excellent account of the CMM, as does Tim Crane's fine book. Andy Clark also provides an extensive introduction to computation and connectionism. The collection edited by Cohen, Johnston, and Plunkett is a valuable resource for thinking about the way in which computational cognitive psychology, and particularly connectionist modelling, impacts on cognitive neuropsychology. The relevance (or not) of neuroimaging to cognitive psychology is discussed in a Forum in the journal Cortex (2006), with the target paper by Max Coltheart.

Block, N. (1995) 'The mind as the software of the brain', in Smith, E.E. and Osherson, D.N. (eds.) *An Invitation to Cognitive Science, Volume 3: Thinking*, Cambridge, MA, MIT Press. Ned Block's classic paper explains the computational model of the mind in detail, and includes informative discussions of machine intelligence, variable physical realization, the language of thought, explanatory levels, and reduction.

Clark, A. (2001) *Mindware: An Introduction to the Philosophy of Cognitive Science*, Oxford, Oxford University Press. Andy Clark's engaging and exciting book begins with 'mindware as software', continues with classical symbol manipulation and connectionism, and introduces further topics, including robots, dynamical systems, and embodied and embedded cognition.

Cohen, G., Johnston, R.A., and Plunkett, K. (eds.) (2000) *Exploring Cognition: Damaged Brains and Neural Networks. Readings in Cognitive Neuropsychology and Connectionist*

Modelling, Hove, Psychology Press. This collection, with an excellent introductory chapter by Robert Johnston and Nick Braisby, includes papers by Coltheart and colleagues (1993), Juola and Plunkett (2000), and Marslen-Wilson and Tyler (1998), along with a wealth of additional resources.

Coltheart, M. (2006b) 'What has functional neuroimaging told us about the mind (so far)?' *Cortex*, vol.42, pp.323–31. In his position paper, Max Coltheart argues that 'no functional neuroimaging research to date has yielded data that can be used to distinguish between competing psychological theories'; this is followed by comments and a response.

Crane, T. (2003) *The Mechanical Mind: A Philosophical Introduction to Minds, Machines and Mental Representations*, 2nd edition, London, Routledge. Tim Crane's excellent book explains representation, the computational model of the mind, and connectionism; it includes brief discussions of systematicity and modularity and concludes with a chapter on consciousness.

REFERENCES

Barkow, J.H., Cosmides, L., and Tooby, J. (eds.) (1992) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, Oxford, Oxford University Press.

Baron-Cohen, S., Tager-Flusberg, H., and Cohen, D. (eds.) (1999) *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*, 2nd edition, Oxford, Oxford University Press.

Bauer, R.M. (1984) 'Autonomic recognition of names and faces in prosopagnosia: a neuropsychological application of the guilty knowledge test', *Neuropsychologia*, vol.22, pp.457–69.

Block, N. (1995) 'The mind as the software of the brain', in Smith, E.E. and Osherson, D.N. (eds.) *An Invitation to Cognitive Science*: *Volume 3 – Thinking*, Cambridge, MA, MIT Press.

Bruce, V. and Young, A. (1986) 'Understanding face recognition', *British Journal of Psychology*, vol.77, pp.305–27.

Caramazza, A. (1986) 'On drawing inferences about the structure of normal cognitive systems from the analysis of patterns of impaired performance: the case for single-patient studies', *Brain and Cognition*, vol.5, pp.41–66.

Carruthers, P. (2006) *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*, Oxford, Oxford University Press.

Chomsky, N. (2000) *New Horizons in the Study of Language and Mind*, Cambridge, Cambridge University Press.

Chomsky, N. (2002) *On Nature and Language*, Cambridge, Cambridge University Press.

Chomsky, N. (2003) 'Replies', in Antony, L.M. and Hornstein, N. (eds.) *Chomsky and his Critics*, Oxford, Blackwell.

Churchland, P.M. (1989) *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA, MIT Press.

Churchland, P.M. and Churchland, P.S. (1996) 'Replies from the Churchlands', in McCauley, R.N. (ed.) *The Churchlands and their Critics*, Oxford, Blackwell Publishers.

Churchland, P.M. and Churchland, P.S. (1998) 'Intertheoretic reduction: a neuroscientist's field guide', in Churchland, P.M. and Churchland, P.S. *On the Contrary: Critical Essays*, 1987–1997, Cambridge, MA, MIT Press.

Churchland, P.S. (1986) *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, Cambridge, MA, MIT Press.

Coltheart, M. (1999) 'Modularity and cognition', *Trends in Cognitive Sciences*, vol.3, pp.115–20.

Coltheart, M. (2001) 'Assumptions and methods in cognitive neuropsychology', in Rapp, B. (ed.) *The Handbook of Cognitive Neuropsychology: What Deficits Reveal About the Human Mind*, Hove, Psychology Press.

Coltheart, M. (2006a) 'Acquired dyslexias and the computational modelling of reading', in Coltheart, M. and Caramazza, A. (eds.) *Cognitive Neuropsychology Twenty Years On*, Hove, Psychology Press.

Coltheart, M. (2006b) 'What has functional neuroimaging told us about the mind (so far)?' *Cortex*, vol.42, pp.323–31.

Coltheart, M. and Davies, M. (2003) 'Inference and explanation in cognitive neuropsychology' *Cortex*, vol.39, pp.188–91.

Coltheart, M. and Langdon, R. (1998) 'Autism, modularity and levels of explanation in cognitive science', *Mind & Language*, vol.13, pp.138–52.

Coltheart, M., Curtis, B., Atkins, P., and Haller, M. (1993) 'Models of reading aloud: dual-route and parallel-distributed-processing approaches', *Psychological Review*, vol.100, pp.589–608. (Reprinted in Cohen, G., Johnston, R.A., and Plunkett, K. (eds.) *Exploring Cognition: Damaged Brains and Neural Networks. Readings in Cognitive Neuropsychology and Connectionist Modelling*, Hove, Psychology Press, 2000.)

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001) 'DRC: a dual route cascaded model of visual word recognition and reading aloud', *Psychological Review*, vol.108, pp.204–56.

Cosmides, L. and Tooby, J. (1994) 'Origins of domain specificity: the evolution of functional organization', in Hirschfeld, L.A. and Gelman, S.A. (eds.) *Mapping the Mind: Domain*

Specificity in Cognition and Culture, Cambridge, Cambridge University Press.

Davies, M. (1995) 'Two notions of implicit rules', in Tomberlin, J. (ed.) *Philosophical Perspectives, Volume 9: AI, Connectionism and Philosophical Psychology*, Oxford, Blackwell.

Davies, M. (2010) 'Double dissociation: understanding its role in cognitive neuropsychology', *Mind & Language*, vol.25, pp.500–40.

Davies, M. and Humphreys, G.W. (eds.) (1993) *Consciousness: Psychological and Philosophical Essays*, Oxford, Blackwell Publishers.

Davies, M. and Stone, T. (eds.) (1995) *Folk Psychology: The Theory of Mind Debate*, Oxford, Blackwell Publishers.

Dennett, D.C. (1984) 'Cognitive wheels: the frame problem of AI', in Hookway, C. (ed.) *Minds, Machines, and Evolution: Philosophical Studies*, Cambridge, Cambridge University Press. (Reprinted in Boden, M.A. (ed.) *The Philosophy of Artificial Intelligence*, Oxford, Oxford University Press, 1990.)

Ellis, A.W. and Young, A.W. (1998) *Cognitive Neuropsychology: A Textbook with Readings*, Hove, Psychology Press.

Fodor, J.A. (1975) *The Language of Thought*, Cambridge, MA, Harvard University Press.

Fodor, J.A. (1981) 'Some notes on what linguistics is about', in Block, N. (ed.) *Readings in the Philosophy of Psychology, Volume II*, Cambridge, MA, Harvard University Press. (Reprinted in Katz, J. (ed.) *The Philosophy of Linguistics*, Oxford, Oxford University Press, 1985.)

Fodor, J.A. (1983) *The Modularity of Mind*, Cambridge, MA, MIT Press.

Fodor, J.A. (1985) 'Précis of *The Modularity of Mind*', *Behavioral and Brain Sciences*, vol.8, no.1, pp.1–42. (Reprinted in Fodor, 1990, pp.195–206.)

Fodor, J.A. (1987) *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA, MIT Press.

Fodor, J.A. (1990) *A Theory of Content and Other Essays*, Cambridge, MA, MIT Press.

Fodor, J.A. (1998) *In Critical Condition*, Cambridge, MA, MIT Press.

Fodor, J.A. (2000) *The Mind Doesn't Work that Way: The Scope and Limits of Computational Psychology*, Cambridge, MA, MIT Press.

Fodor, J.A. (2005) 'Reply to Steven Pinker "So how *does* the mind work?"' *Mind & Language*, vol.20, pp.25–32.

Fodor, J.A. (2008) *LOT 2: The Language of Thought Revisited*, Oxford, Oxford University Press.

Gallistel, C.R. (2001) 'Mental representations, psychology of', in Smelser, N.J. and Baltes, P.B. (eds.) *International Encyclopedia of the Social and Behavioral Sciences*, Amsterdam and New York, Elsevier.

Garson, J. (2010) 'Connectionism', in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* [online] Available from: http://plato.stanford.edu/entries/connectionism/

Gold, I. and Stoljar, D. (1999) 'A neuron doctrine in the philosophy of neuroscience', *Behavioral and Brain Sciences*, vol.22, pp.809–69.

Henson, R.N.A. (2005) 'What can functional neuroimaging tell the experimental psychologist?' *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, vol.58, pp.193–233.

Joanisse, M.F. and Seidenberg, M.S. (1999) 'Impairments in verb morphology after brain injury: a connectionist model', *Proceedings of the National Academy of Science*, vol.96, pp.7592–7.

Joanisse, M.F. and Seidenberg, M.S. (2005) 'Imaging the past: neural activation in frontal and temporal regions during regular and irregular past-tense processing', *Cognitive, Affective and Behavioral Neuroscience*, vol.5, pp.282–96.

Juola, P. and Plunkett, K. (2000) 'Why double dissociations don't mean much', in Cohen, G., Johnston, R.A., and Plunkett, K. (eds.) *Exploring Cognition: Damaged Brains and Neural Networks. Readings in Cognitive Neuropsychology and Connectionist Modelling*, Hove, Psychology Press.

Ladyman, J. (2002) *Understanding Philosophy of Science*, London, Routledge.

Lashley, K.S. (1930) 'Basic neural mechanisms in behavior', *Psychological Review*, vol.37, pp.1–24.

Marr, D. (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, New York, W.H. Freeman and Company. (Reprinted Cambridge, MA, MIT Press, 2010). (Chapter 1, 'The philosophy and the approach', reprinted in Yantis, S. (ed.) *Visual Perception: Essential Readings*, Hove, Psychology Press, 2003.)

Marslen-Wilson, W. and Tyler, L.K. (1987) 'Against modularity', in Garfield, J. (ed.) *Modularity in Knowledge Representation and Natural Language Understanding*, Cambridge, MA, MIT Press.

Marslen-Wilson, W. and Tyler, L.K. (1997) 'Dissociating types of mental computation', *Nature*, vol.387, pp.592–4.

Marslen-Wilson, W. and Tyler, L.K. (1998) 'Rules, representations, and the English past tense', *Trends in Cognitive Sciences*, vol.2, pp.428–35. (Reprinted in Cohen, G., Johnston, R.A., and Plunkett, K. (eds.) *Exploring Cognition: Damaged Brains and Neural Networks. Readings in Cognitive Neuropsychology and Connectionist Modelling*, Hove, Psychology Press, 2000.)

Marslen-Wilson, W., Brown, C.M., and Tyler, L.K. (1988) 'Lexical representations in spoken language comprehension', *Language and Cognitive Processes*, vol.3, pp.1–16.

Marslen-Wilson, W., Tyler, L.K., and Koster, C. (1993) 'Integrative processes in utterance resolution', *Journal of Memory and Language*, vol.32, pp.647–66.

McClelland, J.L. and Patterson, K. (2002a) '"Words or rules" cannot exploit the regularity of exceptions', *Trends in Cognitive Sciences*, vol.6, pp.464–5.

McClelland, J.L. and Patterson, K. (2002b) 'Rules or connections in past-tense inflections: what does the evidence rule out?' *Trends in Cognitive Sciences*, vol.6, pp.465–72.

McClelland, J.L., Rumelhart, D.E., and Hinton, G.E. (1986) 'The appeal of parallel distributed processing', in Rumelhart, D.E. and McClelland, J.L. (eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations*, Cambridge, MA, MIT Press.

Pinker, S. (1997) *How the Mind Works*, New York, W.W. Norton.

Pinker, S. (2005a) 'So how *does* the mind work?' *Mind & Language*, vol.20, pp.1–24.

Pinker, S. (2005b) 'A reply to Jerry Fodor on how the mind works', *Mind & Language*, vol.20, pp.33–8.

Pinker, S. (2006) 'Whatever happened to the past tense debate?' in Baković, E., Ito, J., and McCarthy, J.J. (eds.) *Wondering at the Natural Fecundity of Things: Essays in Honor of Alan Prince*, Santa Cruz, CA, Linguistics Research Center, University of California, Santa Cruz. Available online from:

http://www.escholarship.org/uc/item/0xf9q0n8

Pinker, S. and Prince, A. (1988) 'On language and connectionism; analysis of a parallel distributed processing model of language acquisition', *Cognition*, vol.28, pp.73–193.

Pinker, S. and Ullman, M.T. (2002a) 'The past and future of the past tense', *Trends in Cognitive Sciences*, vol.6, pp.456–63.

Pinker, S. and Ullman, M.T. (2002b) 'Combination and structure, not gradedness, is the issue', *Trends in Cognitive Sciences*, vol.6, pp.463–74.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S., and Patterson, K. (1996) 'Understanding normal and impaired word reading: computational principles in quasi-regular domains', *Psychological Review*, vol.103, pp.56–115.

Plunkett, K. and Bandelow, S. (2006) 'Stochastic approaches to understanding dissociations in inflectional morphology', *Brain and Language*, vol.98, pp.194–209.

Plunkett, K. and Marchman, V. (1993) 'From rote learning to system building: acquiring verb morphology in children and connectionist nets', *Cognition*, vol.48, pp.21–69.

Robbins, P. (2009) 'Modularity of mind', in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* [online] Available from: http://plato.stanford.edu/entries/modularity-mind/

Rumelhart, D.E. and McClelland, J.L. (1986) 'On learning the past tenses of English verbs', in McClelland, J.A. and Rumelhart, D.E. (eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models,* Cambridge, MA, MIT Press.

Searle, J.R. (1992) *The Rediscovery of the Mind*, Cambridge, MA, MIT Press.

Seidenberg, M.S. and McClelland, J.L. (1989) 'A distributed, developmental model of word recognition and naming', *Psychological Review*, vol.96, pp.523–68.

Seidenberg, M.S. and Plaut, D.C. (2006) 'Progress in understanding word reading: data fitting versus theory building', in Andrews, S. (ed.) *From Inkmarks to Ideas: Current Issues in Lexical Processing*, Hove, Psychology Press.

Shallice, T. (1988) *From Neuropsychology to Mental Structure*, Cambridge, Cambridge University Press.

Shanahan, M. (2009) 'The frame problem', in Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* [online] Available from: http://plato.stanford.edu/entries/frame-problem/

Sperber, D. (1994) 'The modularity of thought and the epidemiology of representations', in Hirschfeld, L.A. and Gelman, S.A. (eds.) *Mapping the Mind: Domain Specificity in Cognition and Culture*, Cambridge, Cambridge University Press.

Sperber, D. (1996) *Explaining Culture: A Naturalistic Approach*, Oxford, Blackwell.

Sperber, D. (2002) 'In defense of massive modularity', in Dupoux, I. (ed.) *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*, Cambridge, MA, MIT Press.

Stone, T. and Davies, M. (1999) 'Autonomous psychology and the moderate neuron doctrine', *Behavioral and Brain Sciences*, vol.22, pp.849–50.

Stone, T. and Young, A.W. (1997) 'Delusions and brain injury: the philosophy and psychology of belief', *Mind & Language*, vol.12, pp.327–64.

Swinney, D.A. (1979) 'Lexical access during sentence comprehension: (re)consideration of context effects', *Journal of Verbal Learning and Verbal Behavior*, vol.18, no.5, pp.645–59.

Tyler, L.K. and Marslen-Wilson, W. (1977) 'The on-line effects of semantic context on syntactic processing', *Journal of Verbal Learning and Verbal Behavior*, vol.16, pp.683–92.

Tyler, L.K. and Marslen-Wilson, W. (1982) 'Processing utterances in discourse contexts: on-line resolution of anaphors', *Journal of Semantics*, vol.1, pp.297–314.

Ullman, M.T. (2001) 'A neurocognitive perspective on language: the declarative/procedural model', *Nature Reviews Neuroscience*, vol.2, pp.717–26.

Ullman, M.T., Corkin, S., Coppola, M., Hickol, G., Growdon, J.H., Koroshetz, W.J., and Pinker, S. (1997) 'A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system', *Journal of Cognitive Neuroscience*, vol.9, pp.266–76.

Ullman, M.T., Pancheva, R., Love, T., Lee, E., Swinney, D., and Hickok, G. (2005) 'Neural correlates of lexicon and grammar: evidence from the production, reading, and judgment of inflection in aphasia', *Brain and Language*, vol.93, pp.185–238.

Weiskrantz, L. and Davies, M. (eds.) (2008) *Frontiers of Consciousness: Chichele Lectures*, Oxford, Oxford University Press.

Woollams, A.M., Joanisse, M.F., and Patterson, K. (2009) 'Past-tense generation from form versus meaning: behavioural data and simulation evidence', *Journal of Memory and Language*, vol.61, pp.55–76.

Young, A.W. (1998) *Face and Mind*, Oxford, Oxford University Press.

Visit www.oxfordtextbooks.co.uk/orc/braisby2e/ for online resources accompanying this chapter.

This page intentionally left blank

EPILOGUE

In Chapter 1 of this volume, we attempted to lay out some of the foundations of cognitive psychology, in part by tracing some of its historical antecedents. We hope that, having reached the end of the book, you now find several of the features we described in that first chapter are in much sharper focus.

One fundamental aspect of the mind is its intentional nature – mental states are *about* aspects of the world and the discussions throughout this volume have illustrated this. In Part 1, we saw how the visual system recovers information about perceived objects from the patterns of light that fall on the retina. Part 2 outlined the processes by which the language system recovers information about perceived words from patterns of sound (in audition) or light (in vision). Part 3 was concerned with our memories of particular aspects of (external) events. Even Part 4, concerned with the seemingly inwardly directed activity of thinking, was similarly concerned with the external aspect of mental representations – at least when thinking is deliberative (Type 2). Thinking about problems, choices, and arguments involves recovering true information about such things if we are to arrive at correct solutions, genuine likelihoods, and valid conclusions. And in Part 5 we saw how cognition can be examined via brain activity and how cognition ceases truly to reflect the way the world is when brain organization is disrupted.

Another aspect of the mind that has become clear throughout this volume is its fractionation. In part, the strategy of isolating particular mental faculties from one another is a methodological one, pursued by cognitive psychologists in order to facilitate systematic study. But cognitive psychologists also believe that the mind *is* fractionated, containing multiple interacting components that, acting in concert, give rise to what appears to us to be a unified mind. The different parts of this book also reflect the fractionation – we consider visual perception in isolation from language comprehension partly for reasons of methodological convenience, but partly because we believe that visual perception and language comprehension call on unique cognitive processes.

The fractionation of the mind has been elevated to the status of a philosophical thesis – the so-called modularity of mind – that was the subject of explicit discussion in Chapter 20, and also much (sometimes implicit) discussion elsewhere, for example in Chapter 19. The modularity thesis distinguishes input systems, such as visual perception and the early stages of language comprehension, from central systems, such as reasoning, judgement, and arguably certain aspects of categorization. The extent to which mental faculties reflect distinct processing modules adds further justification to studying them in isolation from one another.

If the mind can be characterized in terms of modularity, the question arises as to what kind of modules it possesses. Most of the chapters in this book have shown that certain cognitive processes are subject to top-down influence, and the influence of quite general knowledge. While such influences do not necessarily undermine a strict interpretation of modularity, one in which processes are informationally encapsulated, they nevertheless suggest that alternative understandings of modularity should be considered (see the discussion of modularity in Chapter 20). Perhaps, for instance, there is a sense of modularity in which modules can influence one another, or can influence one another to a certain degree. The exact nature of the mind's modularity remains an open question – that it has some kind of modular structure seems to be a claim on which many different theorists agree.

In Chapter 20 we also saw an argument that our understanding of central systems is likely always to be poor. If central systems such as reasoning are sufficiently flexible as to be able to call on any aspect of general knowledge, if they are so informationally *un*encapsulated, the argument goes, then it may not be possible to delineate them in a precise or meaningful way. However, much of this book suggests the opposite. Where we have seen the influences of general knowledge on systems such as reasoning, the result has often been an improved understanding of the processes involved. Even straying from the prime candidates for modularity, perceptual systems and the early stages of language comprehension, researchers have succeeded in identifying systematic influences on processing. To take just a few examples, in reasoning, cognitive psychologists have developed theories to account for conditional inference, and have succeeded in devising tasks that evaluate these. Researchers are shedding light on the nature of consciousness, for so long regarded as such a mysterious subject that a scientific approach to its study should be avoided. Cognitive psychologists have also begun to shed light on the influence that emotion exerts upon such cognitive processes as attention, memory, and language comprehension. The evaluation of this research may yet be incomplete, but the picture the research suggests is very different from the negative one offered by the modularity thesis.

The chapters in this book have also illuminated the preferred kinds of explanation of cognitive processes. Chapter 1 outlined the importance of building models in the development of scientific theories and, throughout the book, more or less detailed models of different cognitive processes have been considered. Generally, researchers have developed two kinds of computational model of cognitive processes – symbolic and connectionist. These offer very different kinds of understanding of cognition, and two different styles of explanation. According to symbolic models, cognition involves the rule-based manipulation of (neurally realized) symbols and symbol expressions. In connectionism, cognition is seen as involving the formation of a stable pattern of activation across processing units that, in general, are not symbolic.

As this volume has revealed, there is much debate about the merits of these two styles of computational modelling. Each has arguments in its favour. Symbolic models have advantages in modelling rule-based systems, as language appears to be (though this too is debated). Connectionist models have the advantage of modelling cognition in terms of processes and units that are superficially similar to processes and structures in the brain.

The debate over symbolic and connectionist models hints at a broader tension concerning the kind of explanation at play in cognitive psychology. In Chapter 1, we outlined Marr's three levels of explanation, suggesting that cognitive psychology was concerned more with levels 1 and 2, the computational and algorithmic levels, than with level 3, the hardware or implementation level (see also the discussion of this issue in Chapter 20). Throughout this book, however, we have seen evidence of the close relation between cognitive processes and underlying physiological processes. Indeed, some have questioned whether these are more closely linked than Marr's analysis suggests. For example, connectionists try to ensure that the processes involved in computational models are consistent with what we know of neural processes.

Evidence from people who have experienced brain damage provides valuable evidence of the nature of cognitive processes. If cognition is truly fractionated as researchers believe, then it ought to be possible for certain mental faculties to be impaired through damage whilst leaving others intact. Such dissociations have been reliably observed in many different areas of cognition, and have helped researchers to build, evaluate, and refine ever more complex models of cognitive processes.

By providing even more direct evidence of the nature of neural activity, neuroimaging techniques have enriched our understanding of cognitive processing. Whilst participants perform tasks, images can be generated that indicate regions of the brain with greatest neural activity. If researchers have a prior understanding of the function of those brain areas, they can use this evidence to infer what kinds of information and information processing are involved in different tasks. Of course, it is very difficult indeed to understand precisely the functions of particular brain areas, and so the interpretation of neuroimaging evidence remains hotly debated. However, there seems little doubt that the next phase in the evolution of cognitive psychology will see closer integration between work in the neurosciences and research adopting a more traditional behavioural or experimental approach to the study of cognition. Already we are beginning to see the value of such integration in our understanding of the cognitive bases of psychiatric disorders (see Chapter 15). The intertwining of evidence from neuroimaging, neuropsychological, and behavioural studies is set to become even closer, and shed light on the fine detail of cognition, as in the case of category specificity (see Chapters 1 and 5), to take just one example.

These are just some of the main themes to have emerged from the chapters in this volume. At heart, such themes are all deeply concerned with the fundamental nature of cognition, and how the debates they engender ultimately turn out is likely to have profound implications for our understanding of the mind.

Regardless of the intellectual excitement of debates and disagreements, it would be wrong to characterize cognitive psychology solely in these terms. The chapters in this book also document the very considerable progress researchers have made in establishing a shared understanding of the structure and time course of cognitive processes. It is perhaps this combination of substantial progress on the detail of cognitive processes with profoundly significant debates about fundamental features of the mind that makes cognitive psychology such an exciting subject. We hope the chapters of this book have given you a rich understanding of cognitive psychology and, above all, have enabled you to share this sense of excitement and enthusiasm.

This page intentionally left blank

GLOSSARY

2½D sketch The third stage in Marr's theory of visual perception, in which a description is formed of how the surfaces in view relate to one another and to the observer.

3D object-centred description The fourth stage in Marr's theory of visual perception, in which object descriptions are produced that allow the object to be recognized from any angle (i.e. independent of the viewpoint of the observer).

Abductive inference Inference to the best explanation. In contrast to deductive inference, inductive inference and abductive inference are kinds of non-demonstrative inference: the conclusion goes beyond what is logically guaranteed by the premise or premises. Abductive inference differs from inductive inference in that it appeals to explanatory considerations and not just statistical or probabilistic considerations. In abductive inference, the premises state some evidence and the conclusion is the hypothesis or theory which, if it were true, would provide the best explanation of the evidence.

Acquisition time In fMRI, the time required to take one scan.

Activation PET PET studies used to determine where activity increases in the brain as a function of a particular cognitive task.

Afferent Afferent nerve fibres run 'in towards' the nervous system. They conduct (for example) sensory information, and contrast with efferent fibres.

Affordance Refers to the fact that the visual properties of an object provide cues as to how it can be interacted with and used.

Agnosia Impaired perceptual recognition not due to a deficit of sensory processes. Apperceptive agnosia refers to a deficit of higher level perceptual processes. In these cases, patients may not be able to recognize objects if presented in unusual views or overlapping other stimuli. Associative agnosia refers to a deficit in retrieving semantic information about the stimulus.

Alexia Impairment of reading. The term alexia is mainly used to imply an acquired deficit following brain damage.

Algorithm A clear-cut formula or set of steps for solving a particular problem.

Allophone Allophones are speech sounds that are distinct but that do not change the meaning of a word if interchanged; for example, the 'o' in scone can be pronounced differently depending on regional accent, but the listener is still able to recognize the word scone with little difficulty.

Ambient optic array The structure that is imposed on light when it is reflected by the textured surfaces in the world around us.

Amnesic syndrome This syndrome refers to a relatively pure memory deficit for the learning of new episodic information. It includes a retrograde component extending back months or even some years from the amnesia-inducing trauma, but amnesics retain normal memory for events that occurred prior to this period.

Amygdalae The amygdalae (singular, amygdala) are groups of nuclei found within the medial temporal lobes. They have been shown to play a significant role in emotion.

Analogical problem solving Using an analogy from a semantically distant domain to guide problem solving.

Anaphora Property of referring back to a previous mention (e.g. a man) using either a definite noun phrase (e.g. the man) or pronoun (e.g. he). Literally means 'carrying back'.

Angiograms Maps of blood vessel distribution.

Anosognosia Unawareness of cognitive, sensory, or motor impairment due to brain damage. Patients showing anosognosia tend to deny or grossly underestimate their own disabilities, such as paresis, amnesia, aphasia, or blindness, despite these deficits being evident to other persons.

Anterior In front or before; contrasts with posterior.

Anterograde amnesia Difficulty in learning new information after the brain lesion.

Apraxia Impairment of gesture production not due to motor impairment *per se*.

Arousal A state of being awake and responsive to stimuli. Arousal depends in part on the activity of the autonomic nervous system (ANS).

Articulatory loop (phonological loop) A sub-vocal rehearsal loop.

Articulatory suppression A secondary task (see Dual task) involving the repetition of redundant and irrelevant words such as the-the-the-the.

Associated words Associated words are words that seem to go naturally together, such as cheddar and cheese, or fish and chips. Association between words is often determined by asking people to say or write down the first word that comes into their head when they read a target word.

Attentional blink When someone watching for a target item in a rapidly presented sequence of stimuli detects the target, there follows a brief period when they are unable to identify subsequent items. It is as if their attention has 'closed its eyes' – a blink (see Rapid serial visual presentation).

Attenuation The reduction in intensity of a stimulus. Wearing sunglasses attenuates the light. It has been suggested that neural activity associated with a stimulus, such as speech, is attenuated when attention is directed elsewhere.

Auditory perception The cognitive processes involved in making sense of information received from the ears.

Automatic process A process whose successful conduct has come to require little attention. Examples include behaviours such as driving, and are usually the result of learning, repetition, and practice.

Autonomic nervous system (ANS) A network of nerve fibres throughout the body that transmits signals to the various organs, muscles, and glands. The ANS is divided into two sections. The sympathetic ANS produces effects associated with arousal, while the parasympathetic ANS tends to dominate during periods of rest, and has broadly opposite effects on the body.

Backward masking A technique used in cognitive research whereby the presentation of an initial stimulus (the target) is disrupted by the subsequent presentation of a second stimulus (the mask), often to the extent that a participant will fail to report seeing the target.

Balint's syndrome An extreme form of neglect in which both parietal lobes are damaged, making it hard for a patient to shift attention between objects.

Base rate The information about the broad likelihood of a particular type of event.

Baseline correction In neuroimaging, a process that involves calculating, for each recording channel, the mean signal over a specified baseline interval, prior to stimulus onset, which is then subtracted from the signal at all time points.

Basic emotions A categorical set of discrete emotions, often thought to be of biological origin, and considered to be the most fundamental or important. It is argued that different combinations of basic emotions can produce the full richness of emotion experience.

Between-category distinction A form of object recognition that is based on deciding what category an object belongs to. For example, deciding that an object is a car, rather than which specific car it might be.

Binding problem The problem of how the apparent unity of conscious experience is brought about by the distributed activities of the central nervous system. For example, how the features of different objects in the environment come to be perceived as 'belonging together'.

Blindsight A phenomenon in which people who are perceptually blind in a certain area of their visual field are nevertheless able to demonstrate some response to visual stimuli.

Block designs A kind of experimental design in which trials or sets of trials are grouped together in blocks, each block presenting participants with trials combining particular experimental conditions.

Blood oxygenation Refers to the amount of oxygen in the blood within a region of interest.

BOLD contrast Blood oxygen level dependent or BOLD contrast reflects the degree of oxygenation of the blood.

Borderline items Whose category membership is unclear. Colour categories, for example, have no obvious boundary – it seems impossible to draw a line on the colour spectrum, say, between red and orange. In between these two categories there seem to be shades that are neither unequivocally red nor unequivocally orange, hence our use of phrases such as 'a reddy-orange'.

Bottom-up processing An approach to cognition that begins with information received from the senses, which is analysed in order to make sense of the environment. Bottom-up processing assumes that pieces of information in the input signal are initially processed separately, and that these pieces of information are gradually linked to form larger pieces of information as processing continues. For example, when you hear a sentence you would first process the phonetic features in the input before combining these features in order to generate a string of phonemes; these phonemes would then be combined to form words, with words then being combined to form a complete sentence. In other words, processing begins with the smallest processing unit, and builds to produce more complex representations.

Bounded rationality The idea that in decision making, rationality of individuals is limited by what information they have, the cognitive limitations of their minds, and the limited amount of time they have to make decisions.

Broca's aphasia A language impairment mainly consisting of difficulties with speech associated with minor deficit of verbal comprehension.

Canonical coordinate frame A form of 3D object-centred description proposed by Marr and Nishihara (1978) in which objects are represented within a framework that was about the same shape as the object, so that a carrot would be a cylinder that tapered towards one end.

Capacity Limit on the amount of information that can be stored or processed.

Capgras delusion/syndrome A neuropsychological condition in which the patient believes that someone they know has been replaced by an impostor with the same appearance. This can be part of a psychiatric illness but usually results from brain injury.

Cartesian theatre A derisive phrase coined by philosopher Daniel Dennett to describe a hypothetical location in the mind where consciousness is located.

Categories Talk of concepts presupposes the existence of a corresponding category, that which a concept is about. Researchers often adopt the terminological distinction that the word 'concept' refers to something in the mind and 'category' refers to those things in the world that a concept is about.

Categorization Categorization, according to Bruner *et al.* (1956) is to 'render discriminably different things equivalent, to group objects and events and people around us into classes, and to respond to them in terms of their class membership rather than their uniqueness'. Concepts may therefore be implicated when similar behaviour is shown toward different objects, and different behaviour is shown toward different objects. For example, patting an Alsatian and a poodle, but not a wisteria, would be explained by the first two items but not the third sharing category membership.

Category specific impairment Impairment of knowledge about a specific category of objects (e.g. tools) caused by brain injury.

Central executive Hypothetical entity responsible for the attentional control and coordination of mental operations.

Chaining Forming a sequence of items or actions linked by associations between successive pairs of items.

Chinese Room thought experiment An important thought experiment introduced by John Searle. The experiment consists of Searle in a room, acting as a computer might. By hypothesis, through following a rule book and manipulating cards with Chinese symbols written upon them, he can receive input symbols (questions posed by someone outside the room) and output appropriate symbols in return (answers to the self-same questions). To people outside the room, Searle understands Chinese. However, Searle contends that he lacks such an understanding, and merely slavishly follows the rules that are specified. The thought experiment has been much debated and attacked, but Searle used it to criticize the idea that computers, through syntactic manipulation of symbols alone, could ever be claimed to achieve understanding and intelligence. Thus Searle's argument is a denial of the validity of the Turing test.

Chunk A set of items of information grouped, or chunked, so as to be cognitively processed as a single unit.

Classical view The classical view is a view of concepts that has its roots in classical thought. It is the view that things belong to categories because they possess certain properties in common. Accordingly, if something is a member of a category, then it necessarily possesses the common properties, and if something possesses the common properties this is a sufficient criterion for it to be a member of the category. The view is also synonymous with the claim that concepts provide definitions of their corresponding category.

Coarticulation Coarticulation occurs when two sounds that are adjacent in the speech stream blend together. Coarticulation occurs because you have to prepare to say upcoming phonemes well before they are produced. These preparations will lead to changes in phonemes that are currently being produced. For example, the /d/ phonemes in do and dah sound slightly different because of the following vowel.

Co-evolution of theories Two scientific theories are said to co-evolve when they develop in a relationship of reciprocal influence, with correction, constraint, inspiration, and enrichment operating in both directions. Two theories that co-evolve may retain their own distinctive concepts and explanatory perspectives.

Cognitive appraisals Cognitive interpretations of physiological states of arousal.

Cognitive bias The *selective processing* of pathology-congruent information. In other words, cognitive processes are biased if an individual selectively prioritizes information that reinforces a pathological state.

Cognitive bias modification (CBM) Procedures that experimentally train participants to select for further processing information that is either congruent or incongruent with their pathology.

Cognitive correlates of consciousness Cognitive processes that are (merely) correlated with conscious awareness of stimuli in the environment or in memory, and are not necessarily constitutive of consciousness itself.

Cognitive modelling Cognitive modelling involves building a working model of a cognitive process and then comparing the behaviour of that model against human performance.

Cognitive neuropsychology The study of the effect of brain lesions in order to understand intact and impaired cognition, rather than in order to understand which anatomical area(s) implement particular functions (which is the business of the complementary discipline of neuropsychology).

Coherence Property of connected discourse whereby the sentences relate logically or psychologically to each other. Sentences may be linked together through repeated reference to the same people or things.

Common ground Knowledge that is shared between two people such that they both know that it is shared. This is often important for formulating utterances in such a way that they can definitely be understood by the person you are talking to. For example, if it is common ground that two of you were at the same party last night, you could say 'Did you enjoy it, then?' with *it* meaning the party.

Commonsense theories Murphy and Medin argued that concepts are explanation-based; that there is some explanatory principle or theory that unites the category. They offer the example of someone at a party who jumps into a swimming pool fully clothed. You might categorize this person as being intoxicated. But a similarity-based view cannot explain this because your concept of 'intoxicated' is unlikely to include the property 'jumps into swimming pools fully clothed'. So how might we explain the categorization? Murphy and Medin argue that categorizing the person as 'intoxicated' plays a role in explaining their behaviour, that is, in explaining why they jumped into the swimming pool.

Complex concepts Researchers have tried to explain the meanings of phrases and larger linguistic units in terms of complex concepts, i.e. combinations of lexical concepts. The meaning of the phrase 'red car' would then be explained in terms of the combination of the constituent lexical concepts 'red' and 'car'.

Complex problem solving Solving problems that require extensive domain knowledge.

Complex system A system composed of interconnected parts that as a whole exhibit one or more properties (behaviour being among the possible properties) that are not obvious from the properties of the individual parts.

Compositional structure Compositionality is the determination of the meaning of a complex representation by the meanings of its constituent parts and the way that the parts are put together. When a complex representation exhibits compositionality it is also said to have compositional structure. The compositional structure of natural language sentences allows us to understand sentences that we have never heard before.

Computational model of the mind (CMM) Also known as the computational theory of the mind, the CMM is the proposal that the mind is an information-processing system and that cognition involves the rule-guided processing of structured mental representations. The claims that cognitive processes are rule-guided and that mental representations have compositional structure are both challenged by connectionism.

Computational modelling An explicit account of some process or processes that is implemented as a computer program.

Concept Concepts are general ideas formed in the mind, meaning that concepts apply to every one of a class of things (usually described as a category). For example, a concept of 'cat' is a general idea of all cats.

Conceptual incidental test A test of conceptual implicit memory that requires subjects to employ semantic knowledge to answer questions or respond to a cue.

Conditioned reflex An acquired response that is under the control of (conditional on the occurrence of) a stimulus.

Configural processing (and face recognition) This term has been used in several ways, namely to refer to the processing of the spatial relationships between facial features, to the way facial features interact with one another, and to the processing of the holistic nature of the face, i.e. that it is perceived as a whole pattern.

Connectionist systems Artificial neural networks where representation and processing are intimately related. Connectionism supports the memory systems perspective that memory stores and memory processing are localized in the same part of the brain.

Constructability The extent to which a pertinent description of the target in memory can be constructed.

Constructivist approach An approach to perception that concentrates on the 'top-down' flow of information. Thus, perception is guided by testing 'perceptual hypotheses' formed using previously stored knowledge about the environment.

Content words Words that convey meanings, including nouns, verbs, adjectives, and adverbs, are known as content words. These words have meaning even when presented on their own outside of a sentence.

Context-sensitivity Typicality effects are changed by linguistic context. For example, participants rated the typicality of different farm animals with respect to the category 'animal'. Participants were first presented with a context sentence that emphasized a particular activity. When the context sentence referred to milking, cows were considered to be more typical 'animals' than horses. However, when the context referred to riding, horses were considered more typical 'animals' than cows. Other participants rated metal spoons as more typical than wooden spoons, and small spoons as more typical than large spoons, but large wooden spoons as more typical than small wooden spoons. So, the contribution to typicality made by the values large and small depended on whether one was thinking about metal spoons or wooden spoons.

Continuous/spontaneous EEG A technique that produces electroencephalograms reflecting continuous voltage fluctuations as a function of time, and that can be viewed directly for each electrode during the recording.

Contour generator The edge of an object's silhouette, which Marr (1977) suggested could be used to work out the shape of the object.

Contralateral The opposite side from.

Contralesional Opposite the side of the lesion.

Controlled process In contrast to automatic processes, a controlled process is one that requires for its successful execution a considerable amount of attention or cognitive effort. Such processes may become relatively automatic with extensive practice (e.g. driving behaviours), but in their initial stages they are typically slow and effortful.

Covert face recognition Non-conscious face recognition as measured by autonomic responses, which appears to be mediated by different pathways to overt conscious face recognition.

Cross-modal priming In a cross-modal priming task the prime and target are presented in different modalities. Typically the prime word is spoken, followed by a written target. Participants must then make some kind of decision regarding the visual target word. The critical question is whether the prime word influences processing of the target word; if the meaning of the prime word has been activated then it should influence how the target word is processed.

Cybernetics The interdisciplinary study of the structure of regulatory systems.

Decision tree A decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm.

Declarative knowledge Knowledge about events and facts.

Declarative memory system Our memory for events and facts. The key feature of the declarative memory system is that the subject is consciously aware of the information retrieved.

Definitions The classical view of concepts is that concepts provide *definitions* of their corresponding category.

Depolarization The change in voltage across the membrane of a retinal cell as light strikes it. This can be used to derive the intensity of the light and produce a grey level description of the surrounding world.

Depth of processing Craik and Lockhart (1975) presumed that processing followed a fixed sequence of levels, from early perceptual processes, through pattern recognition, to the extraction of meaning. The greater the depth of processing applied to an item, the more elaborate, longer lasting, and stronger were the resulting memory traces, and so the more likely the item was to be remembered.

Depth perception The ability to determine how far away an object is. This can be achieved by comparing the different retinal images from the left and right eyes (known as stereopsis) or by analysing available cues, such as the size of texture elements on the surface of the object.

Derivational morphology Derivational morphology refers to larger modifications to words that may result in a change in grammatical category. For example, the suffix '-ness' can change the adjective 'happy' into a noun: 'happiness'.

Developmental amnesia Impairment in episodic memory despite a preserved memory for facts (semantic memory) and normal language development and literacy levels.

Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM IV) The manual psychiatrists use to categorise psychiatric diagnoses. Published by the American Psychiatric Association, it covers all mental health disorders for both children and adults, it also lists known causes of these disorders, related statistics, and research concerning the best treatments.

Diagnostic value The extent to which memory cues distinguish the target memory record from other records in memory.

Dichoptic A form of visual presentation that delivers a different stimulus to each eye.

Dichotic listening Listening to a different message in each ear (see Shadowing).

Dipoles Neuronal current sources, which may be radial (directed towards or away from the scalp) or tangential (parallel to the scalp).

Direct perception Gibson's (1950) approach to perception, which sees the information present in light as being sufficient to allow a person to move through and interact with the environment without the need for cognitive processing.

Direct retrieval In direct retrieval a cue causes a pattern of activation in autobiographical knowledge (AK) that stabilizes as a specific autobiographical memory and bypasses the stages of generative retrieval.

Discriminability The extent to which the target memory record can be discriminated from all other records in memory.

Distinctiveness The characteristics of a stimulus that makes it stand out from other stimuli. Distinctive processing emphasizes the differences between stimulus items.

Domain specificity An information-processing system is domain specific if its operation is restricted to information about a particular range of properties (e.g. colours) or a particular class of objects (e.g. faces) in the world. A knowledge system can, similarly, be described as domain specific if it contains knowledge only about a restricted range of properties or objects.

Dorsal stream The pathway that projects from the primary visual cortex to the parietal cortex that specializes in information about the position and movement of objects. Sometimes referred to as the 'where system' as it deals with perception for action, or in other words, information about where an object is and how it is moving.

Dot probe task A task in which participants must respond as rapidly as possible to the presentation of a dot (termed a 'probe' because it is used to probe where attention is located).

Double dissociation A pattern of impaired and spared performance in two patients with brain injuries. One patient shows impaired performance on Task 1 but performance within the normal range on Task 2, while a second patient shows the reverse pattern. The two tasks are said to be doubly dissociated. While double dissociation evidence supports models or theories in which separate modules are involved in the two tasks, the inference from evidence to theory is not deductive but, as throughout normal science, abductive.

Down-regulation Down-regulation refers to an individual's ability to reduce an emotional response, and most often applies to negative emotions (e.g. suppressing a bad or unhelpful feeling).

Dual process signal detection theory (DPSDT) Assumes that recognition can enlist two retrieval processes. One retrieval process provides an assessment of the familiarity of items and events, while the second retrieval process is similar to the retrieval process underlying free recall. These two independent retrieval processes are initiated simultaneously.

Dualism An approach to the relationship between mind and body that regards them as belonging to distinct ontological categories. For example, Descartes regarded the mind as a non-physical substance.

Dual-task paradigm (primary task, secondary task) An experimental method in which two tasks are performed simultaneously, with either one (the primary task) being more important than the other (the secondary task) or both being of equal importance.

Dysgraphia Difficulty in writing and spelling.

Dyslexia The term dyslexia is more commonly used to refer to developmental cognitive disorder.

Early onset amnesia Another term for developmental amnesia.

Echoic memory The brief storage of an unattended auditory event (say a spoken word) that enables it still to be 'heard', like an echo, if attention is directed to it sufficiently quickly.

Ecological optics The theory of perception developed by J.J. Gibson (1950) that concentrated on how the environment structures light as it is reflected. Gibson believed that the resultant structure in the light was so rich that no cognitive processing was required to perceive the environment.

EEG (electroencephalography) The detection of the electrical activity of neurons in the brain by means of picking up and amplifying the tiny voltages that reach the scalp.

Efferent Efferent nerve fibres run 'out from' the nervous system; they conduct (for example) motor impulses, and contrast with afferent fibres.

Embodied meaning The idea that the meaning of words can sometimes be directly represented in terms of motor actions that correspond to how you would act out the meaning. For example, the meaning of *kick* might correspond to the motor activation of muscles used in kicking, and the meaning of *give* might correspond to the motor activation of muscles used to hand something to someone.

Emotion family A term used by Paul Ekman to describe sets of emotions that are related to each other.

Emotion regulation The process by which we manage and redirect our emotions.

Encoding Encoding refers to the way in which objects and events in the world are perceived and come to be represented in memory.

Encoding specificity hypothesis Specific retrieval cues facilitate recall only if they and their relation to the target item are stored along with the target item. Successful retrieval of the target item increases with the overlap between the information stored in memory and the information employed at retrieval.

EOG A technique that can be used to monitor eye movements. It involves placing electrodes around the eyes and produces an electrooculogram.

Epiphenomenon (*plural* epiphenomena) The view that consciousness is an epiphenomenon is the view that conscious states are simply by-products of physical (brain) processes and cannot themselves have causal effects.

Episodic The information that defines a particular instance, rather than a general label (such as a letter name; see Semantic). Thus, in addition to recognizing (semantic) the letter P, there is information showing that this instance of the letter is black, upper case, with a comma just to the right of it.

Episodic buffer A limited capacity multi-dimensional store that backs up the central executive and at the same time provides a link with long-term memory.

Episodic memory Episodic memory is the record of our life experiences along with contextual information about these events and occurrences, such as information about where they occurred and the person's thoughts and feelings about the event.

Essential properties Properties such that if an object did not possess them, it would not be that object. The essential properties of birds are properties that all birds necessarily possess; if something doesn't possess them, then it isn't a bird. Essential properties may seem rather like the defining properties of the classical view. However, there is one critical difference. According to psychological essentialism most people will not know what a category's essential properties are, but will still believe that the category has some. We might speculate as to what the essential properties are – perhaps for biological categories they would be genetic properties – but, in general, our beliefs will be vague and may turn out to be incorrect. So psychological essentialism proposes that people's concepts may contain a 'place-holder' for an essence – and the place-holder may even be empty, reflecting a lack of knowledge as to what the essential properties might be.

Essentialism Psychological essentialism is one attempt at more precisely formulating the view that categorization is influenced by deeper, explanatory principles. Medin (1989) and Medin and Ortony (1989) suggested that people believe that, and act as though, category members have certain essential properties in common. That is, people categorize things according to their beliefs about essential properties. They may also believe that the essential properties constrain a category's more superficial properties. For example, the essential properties of birds might be thought to involve their genetic make-up, properties that would constrain their appearance and behaviour.

Evaluability Extent to which an activity or a programme can be evaluated in a reliable and a credible fashion.

Event-related designs A kind of experimental design used in neuroimaging studies that addresses problems associated with block designs by incorporating randomized presentation of trials and averaging neural activity across trials of the same type.

Event-related fields (ERFs) In MEG, the averaged magnetic signals resulting from and time-locked to a particular cognitive event.

Event-related potentials (ERPs) Connecting electrodes to the scalp can reveal electrical potentials (i.e. voltages) produced by neural activity in the brain. Since particular patterns of potential are produced by specific events, such as a flash of light or a particular sound, they are termed 'event-related'. In EEG, ERPs refer to the averaged electrical signals resulting from and time-locked to a particular cognitive event.

Expert A person who possesses a high degree of skill and knowledge in a given domain.

Explicit learning Explicit learning occurs when the learner is fully conscious of what is being learned, and has a strong intention to learn. For example, grammatical rules in a non-native language are typically acquired both consciously and intentionally by second language learners.

Explicit memory With an explicit memory test, it is clear to subjects that they need to remember past events to complete the test. Free recall, cued recall, and recognition tests are all examples of explicit memory tests.

Extended 10/20 system A system for identifying or referencing electrode locations defined with respect to fractions of the distance between nasion and inion (i.e. root of the nose to the 'little bump on the back of your head', or 'front-back') and the pre-auricular points (the little skin lobules in front of your ear canals, or 'left-right', also known as left pre-auricular or LPA and right pre-auricular or RPA).

External representation Knowledge and structure in the environment that may support learning and problem solving, e.g. an abacus, a list written on paper, colour or size.

Familiarity Another term used to indicate strength of evidence.

Feature integration theory Triesman's theory that the separate features of a target stimulus, such as its colour and shape, have to be integrated in order to identify the item. This integration is a serial process (see Serial and Parallel).

Feature recognition theories Theories of object recognition designed to overcome the problems with template matching in which objects are stored in memory and recognized through analysing key characteristics. The Pandemonium system was based on feature recognition and successfully adapted into a computer program that was able to recognize alphanumeric characters.

Feedback When an event is part of a chain of cause-and-effect that forms a circuit or loop, then the event is said to 'feed back' into itself.

Feedback system A process based on monitoring information from sensory receptors in order to moderate the movement of the part of the body where the receptors are located. For example, information from touch receptors on the fingers is used to guide how much pressure to apply when picking up an object.

Filter A system that accepts a mixture of stimuli, but passes on only the attended stimulus while excluding others. This contrasts with the theory that unattended information is only attenuated, not excluded. (See Attenuation.)

Filtering In neuroimaging, a process applied to a data set to remove any frequencies that are not part of the signal of interest.

Fixations In eye movement research a fixation is a period during which the eye remains stable, fixating at the same point. Typically fixations last between 200 and 250 ms.

fMRI Functional magnetic resonance imaging or fMRI is a technique used by psychologists to measure how the concentration of oxygen in the blood changes in different parts of the brain as cognitive state changes.

Focus Linguistic emphasis placed on certain parts of an utterance, either through stressing the speech or through using different syntactic structures.

Fovea An indented area of the retina that contains the highest density of cones and is responsible for the perception of fine detail that is necessary for tasks such as reading.

Frame problem In artificial intelligence, the frame problem is the problem of formally specifying the situation after an action that brings about change. This is problematic because there is a potentially open-ended set of things that might, in principle, have been changed and that have to be specified as not having been changed by the action. In robotics, the problem arises in designing a robot that needs to update its own database to take account of changes produced by its own actions as it performs a task. It might be suggested that this updating would be more tractable if it were restricted to information that was marked as relevant to the task. But relevance depends on context, and what is or is not relevant may itself have been changed by the robot's action.

Framing effect This describes cognitive biases such that presenting the same option in different formats can alter people's decisions. Specifically, individuals have a tendency to select inconsistent choices, depending on whether the question is framed to concentrate on losses or gains.

Free recall Refers to an explicit memory task where the subject is allowed to recall the items of a list in any order.

Function words Function words are grammatical words such as *we* and *on* that often have little meaning on their own, but are important in expressing grammatical relationships between other words in a sentence.

Functional dissociation This occurs when one has identified a brain region that is uniquely influenced by some experimental manipulations but not by others.

Functional fixity A tendency to represent objects in terms of their normal uses, which can block insight when an unusual use is required.

Functionalism The idea that the mind and mental states are constituted solely by their functional role, not by their physical composition. According to functionalism, it is possible for the same function to be implemented in different hardware (e.g. brains and digital computers). For a functionalist, the fact that digital computers and brains are physically different is no argument against them having similar properties (e.g. intelligence).

Functionalist psychology A general psychological philosophy that considers mental life and behaviour in terms of active adaptation to the person's environment. As such, it provides the general basis for developing psychological theories not readily testable by controlled experiments and for applied psychology.

Fuzzy categories Categories that seem to have genuine borderline cases (see also Borderlines). For these, similarity to a prototype might provide the best explanation of category membership, since there is no prospect of defining these categories, and people do not seem to have relevant commonsense theories (e.g. a theory of the deeper causal principles by which red things come to appear red).

Galvanic skin response (GSR) The change in electrical resistance of the skin in response to an emotion-provoking stimulus. The change results from changing levels of sweat in the sweat glands.

Gamma oscillations In neuroimaging, a signal in the frequency range above 30 Hz (though definitions differ slightly), which has been associated with complex stimulus processing such as Gestalt perception.

Gaussian blurring A technique based on averaging the intensity values of an image in circular regions such that the values at the centre of the circle are weighted more than those at the edges in a way identical to a normal (or Gaussian) distribution, which results in blurring the image. Blurring images is an important way of removing random fluctuations and identifying the edges of objects.

General events A variety of autobiographical knowledge structures such as single events (e.g. the day we went to South Wales), repeated events (e.g. Friday afternoons at the pub), and extended events (e.g. the conference in Sydney).

Generalized cone Any 3D shape that has a cross-section of a consistent shape throughout its length. The cross-section can vary in size, but not in shape.

Generative retrieval Generative retrieval occurs when remembering is intentional and the knowledge base is iteratively sampled as a memory is effortfully constructed.

Geon Short form of 'geometric ion' – a set of basic 3D shapes that are used to represent objects in Biederman's theory of recognition.

Gerstmann's syndrome This encompasses four deficits: finger agnosia, agraphia, acalculia, and left-right disorientation. The association of these deficits is not due to damage of similar cognitive processes, but to a common brain lesion in the left parietal lobe.

Gestalt movement An approach to perception dominant in the early 20th century that concentrated on how the parts of an image were organized and held that the 'whole is greater than the sum of its parts'.

Gestaltists A school of psychology originating in Germany and Austria, which claims that all experience consists of 'gestalten', and that the response of an organism to a situation is a complete and unanalysable whole rather than a sum of the responses to specific elements in the situation.

Global workspace theory A theory developed by Bernard Baars in which consciousness is seen as a 'global workspace', a means of bringing together the products of processing from widely distributed modules.

Goal hierarchy A hierarchy of currently active goals and self-conceptions through which it is hypothesized that current experience is encoded and in which autobiographical memories are constructed.

Grey level description The first stage in Marr's theory of visual perception in which the intensity of light is measured at each point in the retinal image.

Greyscale or grey level An image comprised solely from shades of grey ranging from white to black. Such an image can be produced by analysing the intensity of light.

Habituation A decreased neural (psychological or behavioural) response to a stimulus after repeated exposure to that stimulus.

Haemodynamic lag The lag of the order of 4–5 seconds after experimental conditions change before a corresponding change is seen in BOLD contrast.

Haemodynamic response The change in BOLD contrast that occurs with a change in neural activity in a particular brain region.

Haptic information Essentially 'touch' information. The information about objects and surfaces gained from combining kinesthesis with cues from touch receptors.

Haptic perception The cognitive processes involved in making sense of information received through touch.

Head-up display (HUD) A transparent screen placed between an aircraft pilot and the windshield, in which vital information, such as altitude and heading, can be reflected. Being transparent, it is possible to look through the HUD to the outside world, this arrangement eliminating the need to look into the cockpit to read instruments.

Heuristic A shortcut strategy for tackling a class of problems that often but not always yields a solution.

Higher order thought theory A theory (or family of theories) concerning consciousness, in which it is argued that information is conscious only when it is the subject of a (higher order) thought. For example, the experience of a traffic light as red may become conscious only when it is the subject of another thought (I believe that I perceive the traffic light to be red).

Hill-climbing An example general heuristic, which involves taking the immediately possible action that brings the solver closest to the goal. By analogy with climbing a hill in fog by trying out one step in each of four main directions and choosing the one that leads to the highest ground.

Homographs Words that have dual meanings, despite having the same spelling, such as 'batter', 'punch', and 'stalk'.

Homophones Homophones are words that are pronounced in the same way but differ in meaning. Sometimes homophones both sound the same and are spelled the same (e.g. bank – money bank vs river bank) and sometimes they are pronounced the same but spelled differently (e.g. pair vs pear or pane vs pain).

Homunculus A person in the head that can be used to give the misleading impression that we have an explanation for the actions of the owner of the head, when actually the problem of explanation is being avoided.

Horizon ratio relation An example of what Gibson termed invariant information. Refers to the fact that the ratio of how much of an object is above the horizon to how much is below remains constant regardless of how far away the object is.

Iconic memory A system that enables the *experienced* duration of a brief visual stimulus to be longer than the actual event; the visual equivalent of echoic memory.

Impasse A point in problem solving beyond which no further progress appears possible.

Implicit learning Implicit learning occurs when the learner is unaware of what is being learned, and does not necessarily intend to learn. As a result, implicit knowledge can be difficult to describe to others. For example, children typically learn their first language without explicitly being taught grammatical rules, and may later find it difficult to explain these grammatical rules to a non-native speaker.

Implicit memory In tests of implicit memory, performance is influenced by previous experience without the subject intentionally remembering the events from the past or being aware of the relevance of the past events. In other words, implicit memory is supposed to be both unintentional and unconscious.

Indirect speech acts Saying something indirectly to mean something else. Typically used for polite requests (e.g. 'Can you pass the salt?' meaning 'Pass the salt.').

Inflectional morphology A minor modification to a word, for example marking pluralization or tense, is known as inflectional morphology.

Information processing A paradigm that sees cognition as being essentially computational in nature, with the mind being the software and the brain being the hardware.

Informational encapsulation An information-processing system within a creature or larger system is informationally encapsulated if it does not have access to all the information that is represented in the creature or larger system. Specifically, a perceptual system is informationally encapsulated if it does not have access to all of the creature's or larger system's background knowledge. The standard example of informational encapsulation is provided by the persistence of a perceptual illusion even when you know that it is an illusion.

Inner loop monitoring Monitoring what you are about to say before you actually say it.

Insight Restructuring how a problem is represented so that the solution becomes immediately obvious.

Integration A process that can occur in backward masking, where a very brief visual target is followed quickly by a second stimulus. The visual system treats the two events as one, integrating the two sets of information and making the target indecipherable.

Interactive alignment Process whereby interactants come to think the same as each other (i.e. have aligned mental states) as a result of speaking the same way (i.e. using the same words, meanings, and grammatical constructions).

Internal representation Knowledge and structure in memory, as propositions, productions, schemas, neural networks, or other forms.

Interruption A process leading to backward masking. When a second visual stimulus follows a target stimulus, with a delay between them that is a little longer than is required to produce integration, then processing of the target is interrupted before it can be identified.

Intransitivity A source of difficulties for the classical view of concepts has been the observation of intransitivity in categorization judgements. Transitivity is observed with many relationships, e.g. the relation 'taller than' is transitive because if 'A is taller than B' and 'B is taller than C', then it simply follows that 'A is taller than C'. The relationship is 'transitive' because the last statement follows from the first two.

Introspection The self-observation and reporting of conscious inner thoughts, feelings, perceptions, and sensations.

Invariants Higher-order features present in the structure of light that remain constant despite movement or other changes. Gibson believed that they provided unambiguous information about the environment and were a critical part of perception. Examples of invariants include the horizon ratio relation and texture gradient.

Inverse problem The estimation of the sources in the brain that generate the electrical and magnetic activity recorded by EEG or MEG is usually called the inverse problem (also known as source estimation).

Inversion effect (and face recognition) The finding that inverting a photograph of a face disrupts recognition more than inverting an image of an object.

Iowa gambling task A task associated with Antonio Damasio, in which participants are presented with four decks of playing cards. Selecting cards from the decks leads either to gains or losses. The decks are crafted so that some lead to modest gains and others to losses. These differences are not at all obvious, yet after playing for a while participants tend to make selections from the two good decks, even though they are generally unable consciously to identify which decks are advantageous.

Ipsilateral The same side as.

Ipsilesional The same as the lesion.

Isomorphic Sharing a common, underlying structure.

Isotropic A cognitive process is said to be isotropic if there is no *a priori* way of putting a boundary around the knowledge or information that may be relevant to the process. Confirmation of a scientific hypothesis is isotropic because there is no way of circumscribing the evidence that may be relevant.

Item specific processing A type of distinctive encoding that binds the components of the processes underlying the mental representation, enhancing their joint operation and coherence (leading to integration).

Kinaesthetic Referring to kinaesthesia, the sensation produced by movements of the body that allows you to know where the different parts of your body are located at any moment.

Kinesthesis Information about the location of our limbs or other body parts, derived from receptors within muscles and joints.

Know experience Knowing that a test item appeared at study, but not having any conscious recollection of it.

Lateral Surface of a structure situated towards the side of the body, such as the outer side of each brain hemisphere; contrasts with medial.

Lesion Structural alteration of bodily tissue brought about by damage or disease.

Lexical concepts A lexical concept is a concept that corresponds to a single word, and is thought to represent our understanding of the word's meaning. For example, the concept for 'cat' is thought to represent people's understanding of the meaning of the word 'cat'. Lexical concepts are presumed to be stored in what is called the mental lexicon, from which it is thought lexical concepts are retrieved during the process of understanding language.

Lexical decision task A task in which the participant must decide as quickly as possible whether a group of letters spells a word or not.

Lexical effects The term 'lexical effects' simply refers to the effect that word-level information has on lower levels of processing. For example, a sound that is midway between /k/ and /g/ is more likely to be interpreted as a /k/ in the context of '-iss', but as /g/ in the context of '-ift' – our knowledge about what constitutes a real word affects lower-level phonetic and phonological processing.

Lifespan retrieval curve A curve plotting numbers of memories against age at encoding of the remembered experiences.

Lifetime periods Lifetime periods effectively encapsulate a period in memory and in so doing may provide ways in which access to autobiographical knowledge can be limited, channelled, or directed. Like general events, they may contain representations of locations, others, activities, feelings, and goals common to the period they represent.

Long-term memory Long-term memory contains all that we know and all that makes us who we are.

Low frequency words Words that occur only rarely in a language (e.g. tavern).

Magnocellular pathway A distinct 'stream' of relatively large retinal ganglion cells that connect the retinal cells to the primary visual cortex.

Marr and Hildreth's algorithm An algorithm important in Marr's primal sketch stage that uses Gaussian blurring to remove random fluctuations and identify the edges of objects.

Masking A technique used in cognitive research whereby the presentation of one stimulus (the mask) disrupts the presentation of a second stimulus (the target), often to the extent that a participant will fail to report seeing the second stimulus.

Materialism The thesis that the only thing that exists is matter, and therefore that all things in existence are composed of matter. The thesis is in opposition to Cartesian dualism.

Means–ends analysis Analysing a problem into goals, subgoals, and sub-sub-goals, and so on until sub-goals are identified that can be immediately reached by some action.

Medial surface Surface of a structure situated towards the mid-line of the body, such as the inner side of each brain hemisphere; contrasts with lateral.

MEG MEG is a technique that produces the magnetoencephalogram, a record of the magnetic fields above the scalp surface.

Memory store A memory store is where non-active memory representations are held.

Memory system Includes memory stores and all the processes that operate when memory representations are active.

Mental imagery (visual imagery) Images formed on the basis of memory or imagination rather than through stimulation of the senses.

Mental lexicon This is the store of information about all of the words that we know, including what the word sounds like, what it means, and how it is spelled.

Mental model An internal representation having – in some abstract sense – the same structure as the aspect of external reality that it represents.

Mere exposure effect The finding that people tend to prefer items to which they have previously been exposed over comparable novel ones.

Metrical foot Metrical foot refers to the rhythm within a speech stream. In English one unit of metrical foot is typically composed of one stressed (strong) syllable, followed by one or more unstressed (weak) syllables.

Mini–mental state examination (MMSE) A short test used to screen for cognitive impairment. A maximum score of 30 points is possible. People with Alzheimer's disease generally score 26 points or less.

Mind–body problem The mind–body problem is the problem that appears to be posed by the fact that minds and mental phenomena seem to be qualitatively distinct from physical brains and bodies. The problem therefore arises as to how the nature of the mind might be explained in terms of physical properties of the brain/body.

Mnemonic A technique used to aid remembering, such as rhyming or the use of visual imagery.

Modularity The idea that the mind is composed of distinct processors or processing modules, which are typically conceived of as being domain-specific and informationally encapsulated. There is considerable division of opinion over whether such modules have the remaining characteristics proposed by Jerry Fodor (such as innateness), yet Fodor's characterization remains a powerful statement of modularity.

Mood congruent memory (MCM) A robust phenomenon in which memory is enhanced when the nature of the material to be remembered matches present mood.

Mood dependent memory (MDM) A controversial idea that memory for a particular stimulus or event will be enhanced if mood at encoding matches mood at retrieval.

Morpheme A morpheme is the smallest meaningful unit of language that still has meaning. It is important to note that not all morphemes can stand as words on their own; for example, the past tense ending 'ed' clearly conveys information about tense, but cannot stand on its own.

Morphology Morphology is the study of how words are structured. For example, the word unbreakable can be broken down into three components, 'un', 'break', and 'able'.

Motion parallax The principle that the further an object is from an observer, the less it will appear to move as the observer travels past it.

Multi-store memory model Atkinson and Shiffrin's (1968) 'multi-store' memory model proposed separate sensory registers for each sense modality. See also Short-term memory and Long-term memory.

Necessary/elaborative inferences Text inferences that are either necessary to understanding the meaning or are not necessary but perhaps helpful. For example, when you read 'When Harry drove to London the car kept breaking down', you necessarily infer that the car was the one Harry drove to London in. However, when you read 'Harry jumped off the top of the building', you might elaboratively infer that as a consequence he died, but you might be wrong.

Neglect A shortened form of 'hemispatial neglect'; often used to describe patients, i.e. *neglect patient*. Patients showing neglect syndrome tend to ignore stimuli located on the contralesional (usually the left) side. The terms extrapersonal, personal, or representative neglect are used depending on whether neglect is observed for stimuli located in reaching area, body, or mental images, respectively.

Neighbouring words Words are said to be neighbours if they are spelled similarly (e.g. wade, jade, spade, made).

Neuroimaging data Images of brain function or brain structure.

Neuropsychological data Data that relate the structure and function of the brain to specific psychological processes and behaviours.

Neuropsychology The attempt to draw conclusions about the function of different brain areas by studying the effects of damage to one or more of the brain areas and cerebral networks.

NIRS Near-infrared spectroscopy or NIRS is a technique that uses long-wavelength light to study the oxygenation state of blood in the brain. It is commonly used with infants due to its ease of use and safety.

Node A node is any point in a network where pathways intersect. In cognitive psychology, it means an element representing a concept in a semantic network or relations between nodes.

Non-accidental property Key features of an object that do not vary across different viewpoints. Biederman suggested five properties: curvilinearity; parallelism; cotermination; symmetry; and collinearity.

Non-declarative memory systems Refers to a group of behaviours that are affected by stored information (i.e. memory), but there is no conscious awareness of the information retrieved.

Non-demonstrative inference An inference is demonstrative if it is impossible for the premise or premises of the inference to be true and the conclusion false; that is, if the truth of the premise or premises logically guarantees the truth of the conclusion. A non-demonstrative inference is thus one in which the conclusion is supported by the premises but goes beyond what is logically guaranteed by the premises. Inductive inference (from instances to a generalization) and abductive inference (from evidence to an explanatory hypothesis) are examples of non-demonstrative inferences.

Non-word A pronounceable but meaningless letter string (e.g. mafer).

Norman's dual process approach Uses the characteristics of the dorsal and ventral streams and suggests that they act synergistically so that the dorsal stream is largely concerned with perception for action and the ventral stream is essentially concerned with perception for recognition.

Normative Normative means 'relating to an ideal standard or model'.

Novice A person who is beginning to acquire skill and knowledge in a given domain.

Object-centred description A mental representation of an object that contains information about what it looks like from every angle; in other words, the object is represented fully in three dimensions.

Occluding contour The silhouette of an object that is a key part of Marr's theory of object recognition as it allows the overall shape of an object to be generated.

Operationalism The idea that we do not know the meaning of a concept unless we have a method for determining its presence and measuring its quantity.

Operator A legal move or step used in solving a problem.

Optimal viewing position The optimal viewing position is the point in a word at which that word can be processed fastest and most efficiently if fixated there. This point is usually at the centre or slightly to the left of the centre of the word.

Orthography The orthography of a word is simply the written form of the word, referring to how the word is spelled or represented using a system of symbols (as in Chinese scripts).

Outer loop monitoring Monitoring your own speech on the basis of hearing it.

Own name effect The phenomenon of noticing that one's name has been spoken by another person, even though one is attending closely to what someone else nearby is saying.

Ownership paradigm Paradigm whereby the self-concept is activated in a relatively indirect way through ownership.

Parafovea The parafovea refers to the areas in the retina that surround the fovea. Vision in the parafovea is not as clear as in the fovea, but some information can still be acquired in this region.

Parallel The description of a system that is able to run several different processes simultaneously. (See Serial.)

Parallel activation Parallel activation occurs when multiple words are activated at the same time when speech is heard.

Paresis Partial paralysis, e.g. of one leg.

Parietal One of the lobes of the brain (one in each hemisphere), which is involved in sensory processing, spatial awareness, and object location.

Parsing Parsing involves breaking down a sentence into its component parts in order to obtain the syntactic/grammatical structure of the sentence. For example, the sentence 'The girl saw the cat' can be broken down into three components: the first noun phrase 'the girl'; a verb phrase 'saw'; and a second noun phrase 'the cat'.

Partial information Norman and Bobrow (1975) suggested that less than the final output from a process can allow other processes awaiting this information to begin or may even fully satisfy the requirements of other processes.

Partial report superiority effect When a backward-masked visual stimulus contains several items (e.g. 12 letters), people may be able to report only, say, four of them. If just after the stimulus is removed, a cue directs the participant to a subset of items (e.g. one of three rows), people tend to still report four. Since this works for any cued row it follows that in total all 12 letters were seen, a figure that is *superior* to the four items obtained when attempting a whole report.

Parvocellular pathway A distinct 'stream' of relatively small retinal ganglion cells that connect the retinal cells to the primary visual cortex.

Pattern associator A simple connectionist network with input units and output units, but no hidden units. The two layers of units are fully connected in the sense that each input unit is connected to every output unit. By adjustment of the weights on the connections during training, the network learns to associate patterns of activation across the output units with patterns of activation across the input units.

Perception The cognitive processes of analysing sensory information and providing a basic description of the surrounding environment.

Perceptual classification Another term for 'object recognition', in which an object is categorized. An object in view is compared to descriptions stored in memory and as a result is assigned to the appropriate category.

Perceptual hypotheses A key component of constructivist, or top-down, theories of perception in which previously stored knowledge is used to interpret the world around us by generating a hypothesis about what a particular object or surface might be.

Perceptual incidental test Test of perceptual implicit memory that usually requires subjects to resolve displays that are perceptually impoverished displays.

Perceptual organization A set of principles developed by the Gestalt approach to perception, and also used in Marr's theory of vision, that describe how images can be perceived. Examples of principles include closure, good continuation, proximity, and similarity.

PET Positron emission tomography or PET is another tool, like fMRI, used to measure how blood flow in different parts of the brain changes as a function of cognitive state.

Phoneme A phoneme is a single speech sound, or in other words, the speech equivalent of a single letter. Phonemes are normally denoted by surrounding slash marks, for example, /k/ and /ə/ are the first two phonemes in 'confess'.

Phonemic (phonological) Referring to speech sounds.

Phonetic feature Each phoneme within a language can be classified in terms of a set of phonetic features. These features describe how the phoneme is articulated.

Phonetics Phonetics is the study of the physical properties of speech sounds, such as pitch and amplitude, and how these properties affect both the production and perception of the speech sounds.

Phonological representation The phonological representation of a word specifies what the word sounds like, and can be written as a string of phonemes. For example, the word cat can be represented as /kæt/.

Phonology Phonology is the study of the speech sounds within a language. Phonologists are interested in how speech sounds are combined and organized within a language in order to convey meaning.

Phrase structure grammar Phrase structure grammar can be used to describe the syntactic structure of a sentence. For example, the sentence 'the girl saw the cat' can be described as NP + VP + NP (where NP = noun phrase and VP = verb phrase).

Pneumoencephalogram A low-resolution, poor-contrast x-ray.

Positron emission tomography (PET) A non-invasive technique of brain imaging that detects pairs of gamma rays emitted indirectly by a positron-emitting tracer, introduced into the body on a biologically active molecule.

Posterior Behind or after; contrasts with anterior.

Pragnanz law of An important part of the Gestalt approach to perception and described by Koffka as: 'Of several geometrically possible organizations that one will actually occur which possesses the best, simplest and most stable shape' (Koffka, 1935, p.138).

Primacy The tendency for items at the start of a list to be better remembered than later items.

Primacy debate The debate focused on the question of whether emotion precedes cognition, or whether cognition precedes emotion.

Primal sketch The second stage in Marr's theory of visual perception. It is divided into the raw primal sketch and the full primal sketch. In the raw primal sketch, areas that could potentially correspond to the edges and texture of objects are identified, and then used to generate a description of the outline of any objects in view in the full primal sketch.

Primary memory The content of consciousness supplied by memory rather than perception.

Priming The improvement in the processing of a target stimulus caused by previous or concurrent exposure to a closely related stimulus (the prime), which may also be the target stimulus. At a neural level, this may be reflected in a process of excitation, which makes the neurons that represent the target item more likely to fire if the prime is presented. Priming maybe detected by faster responses, or by the ability to detect an otherwise subthreshold stimulus (e.g. one that is masked). For example, seeing (or hearing) the word 'chair' increases the speed with which you can respond to the word 'table' because these two words are semantically associated.

Primitives A form of symbolic representation used to represent the fundamental properties of a visual image. Marr's theory of vision made use of four primitives: 'edge-segments', 'bars', 'terminations', and 'blobs'.

Problem reduction Transforming the problem into another problem (or sub-problem) for which a solution exists.

Procedural knowledge Skills and habits.

Processing approach Theoretical approach to investigating memory that emphasizes psychological processing.

Productivity A language is said to be productive if speakers (and hearers) of the language are able to use (or comprehend) sentences that they have not previously encountered.

Prominence effect A type of bias that people show when they are asked to provide a valuation of something using a numerical scale. They tend to select prominent values such as 1, 2, 5, 10, 20, 50, 100.

Property-listing (attribute-listing) Property-listing or attribute-listing is a technique in which participants are asked to generate a list of properties or attributes for a concept. Psychologists use the technique to investigate people's concepts. Usually lists from many different participants are compared, to generate a further list of the most frequently mentioned properties. Frequency of mention is often taken to indicate how central each property is to the concept.

Proprioception The processes that allow us to know the locations of our limbs based on combining information from touch receptors with kinaesthetic information.

Prosopagnosia A neuropsychological condition in which the patient is able to recognize objects but not faces, though they may still recognize facial expressions.

Prospect theory A theory that describes decisions between alternatives that involve risk (i.e. alternatives with uncertain outcomes) where the probabilities are known. The model is descriptive: it tries to model real-life choices, rather than optimal decisions.

Protocol analysis A rigorous method for eliciting verbal reports of thought sequences as a valid source of data on participants' thinking as a task is carried out.

Prototype view (of concepts) Whether or not an instance is a category member can be judged depending on how similar it is to a prototype: an instance falls within the category if it achieves a certain criterion of similarity. If an instance is too dissimilar, i.e. it mismatches on too many properties, then it falls outside the category.

Prototype A measure of the central tendency of a category. Sometimes the prototype may correspond to an actual instance, but in general it is like a 'best' category member, formed by statistically aggregating over examples of the category one encounters. Rosch believed that it is a feature of the natural world that certain attributes or properties tend to correlate or cluster together, and it is these natural clusters of correlated attributes that prototypes describe. For example, the prototype for 'bird' might describe the cluster of properties such as having feathers, wings, a beak, and an ability to fly.

Pseudohomophone A pseudohomophone is a non-word that can be pronounced in a way that sounds like a real word (e.g. brane, noyz).

Qualia A term used by philosophers to describe the qualitative, subjective, experiential aspects of consciousness.

Quinean A cognitive process is said to be Quinean if it is sensitive to properties of a system of knowledge or belief as a whole. Confirmation of a scientific hypothesis is Quinean because it is sensitive to properties of whole theories or whole systems of belief, such as simplicity.

Rapid serial visual presentation (RSVP) A procedure for presenting a sequence of several visual stimuli in quick succession, each serving as a mask for the previous item.

Reason-based theory of choice An account of preference formation, according to which an agent's preferences are determined by his or her motivating reasons, together with a 'weighting relation' between different combinations of reasons.

Receiver operating characteristic (ROC) curve Hit rates plotted against false alarm rates for different response criteria. Usually, response criteria are manipulated by asking subjects to rate their confidence that each memory probe is a target.

Recognition memory Refers to an explicit memory task where study items are re-presented together with foils and the subject has to respond 'yes' or 'no' as to whether an item was studied or not.

Recollection A retrieval process underlying recognition, which takes longer to complete than familiarity and benefits from relational processing at encoding. It is also regarded as a threshold process.

Reduced relative clause A reduced relative clause is a relative clause that is not explicitly marked by a relative pronoun. Often these sentences result in ambiguity. For example, when processing the sentence 'The defendant examined by the lawyer remained silent', it is easy to initially assume that the defendant is the person that is going to do the examining when reading the first few words of the sentence. This assumption must be corrected after encountering the phrase 'by the lawyer'.

Reductio ad absurdum Disproof of a proposition by showing that it leads to absurd or untenable conclusions.

Re-entrant pathway/processing Neurological pathway that allows one region of the brain to communicate with another in response to information being sent by that region. In other words, the pathway that allows one region of the brain to respond to another. It has been suggested that this mechanism could allow a perceptual hypothesis to be tested.

Refractory period This is the brief period just after neurons have fired, during which they are unable to fire again. The attentional blink is rather like a refractory period.

Reinforcers A stimulus that, if it is presented soon after a response, increases the relative frequency with which that response is emitted in the future. In other words, a 'reward' that increases the likelihood of a behaviour being repeated.

Relational processing A type of distinctive encoding that establishes connections between different entity representations (leading to elaboration).

Relative clause A relative clause is a phrase that is introduced by a relative pronoun (who, whom, whose, that, which). For example, 'The defendant who was examined by the lawyer remained silent'.

Relative distinctiveness The extent to which a stimulus item differs from other items on a list.

Remember experience Having a conscious recollection that a test item was presented at study.

Reminiscence bump A component of the lifespan retrieval curve when rememberers were aged 10 to 30 years, being distinguished by an increase in recall of memories relative to the period that precedes it and those that follow it.

Repetition priming A technique used in cognitive research whereby the presentation of one stimulus will make recognition of a related stimulus faster and more accurate.

Representation A hypothetical 'internal' cognitive symbol that represents some aspect of external reality.

Requisite decision model This is defined as a model whose form and content are sufficient to solve a particular problem.

Resolution The resolution of a neuroimaging technique indicates how far away, in space (spatial) or time (temporal), two events have to be before they can be distinguished as two separate events.

Response bias A bias at the stage of making a response, which may reveal little or nothing about the nature of prior cognitive processing.

Response criterion Indicates a person's willingness to respond 'yes'. Someone with a liberal response criterion will have a high hit rate and high false-alarm rate.

Retina The back surface of the eye, which contains light-sensitive receptor cells known as rods and cones.

Retinal ganglion cells Connect retinal cells (rods and cones) to the lateral geniculate nucleus, which is in turn connected to the primary visual cortex. The ganglion cell axons leave the eye via the 'blind spot'.

Retrieval The way in which information in memory is obtained for use. Retrieval involves finding, activating, and sometimes further processing pertinent memory representations.

Retrieval intentionality criterion This criterion provides a way to rule out the influence of intentional retrieval on an implicit task.

Retrieval mode In retrieval mode, attention is directed inwards towards internal representations of knowledge, and conscious awareness becomes dominated by these representations.

Retrograde amnesia Difficulty in recalling information encoded before the brain lesion that caused the amnesia.

Saccade Saccades occur when the eyes jump from one location to another; these can be forward saccades or backward (regressive) saccades.

Schema (plural schemata) A mental representation of some aspect of experience, for example routine sequences of actions or mental operations.

Self-cue A cue that indicates self-relevance of a stimulus, for example, a red circle around an object that signifies that the object belongs to self. Self-cues usually elicit self-referential processing.

Self-defining experiences Experiences that are distinctive and unique *for the self*. They are thought to play a crucial role in the final formation of identity and a stable self-system during late adolescence and early adulthood.

Self-defining memories Memories for an individual containing critical knowledge of progress on the attainment of their long-term goals.

Self-defining moments Highly specific events that were 'turning points' for an individual; they are typically found in autobiographical accounts.

Self-reference effect (SRE) The memorial benefit of self-referencing.

Self-referencing Encoding information with reference to oneself.

Self-serving attribution bias A bias in which people attribute good things internally, as something within their control, whereas they attribute bad things externally, to others or to circumstances.

Semantic Relating to meaning and identity, the term is often used to refer to the meaning of a word, or the information that identifies a letter (see Episodic).

Semantic ambiguity Semantic ambiguity occurs when the structure of a sentence remains the same, but the individual words are interpreted in different ways. This often occurs when words have more than one meaning (e.g. he went to the bank).

Semantic features Word meanings can be represented as a set of features of properties that are associated with an item. These features combine to provide the meaning of the word. For example, the word canary has semantic features such as 'has wings', 'can fly', 'is a bird', etc.

Semantic memory Semantic memory is our general knowledge store; it contains all the information underlying our understanding of the world. No personal experience need accompany information retrieved from semantic memory.

Sensation The ability of the sense organs to detect various forms of energy (such as light or sound waves).

Sensitivity Indicates how well a person is able to discriminate targets and distractors. Great sensitivity is indexed by a high hit rate and low false alarm rate.

Sensory neglect Patients who have suffered a lesion in a parietal lobe, particularly in the right hemisphere, display unawareness of items on the opposite side (i.e. generally the left) of their body. This neglect can extend to all senses, but is typically detected visually.

Serial The description of a system that can carry out only one process at a time. (See parallel.)

Servomechanism An automatic device that uses error-sensing negative feedback to correct the performance of a mechanism.

Shadowing The process of repeating everything one hears spoken, following closely behind the speaker's words like a shadow. Shadowing is often used in dichotic listening to ensure that attention is directed to one message stream.

Short-term memory Refers to the retention of information over seconds to minutes.

Signal detection theory (SDT) The SDT originated as a theory about decision making. It assumes that all of the events or items to be discriminated differ only in terms of a continuous (i.e. quantitative), unidimensional measure. People make decisions as to whether test items are old or new based on the items' strength of evidence and the person's own response bias.

Similarity Both classical and prototype views are committed to the idea that similarity explains categorization. For classical theory, instances fall within a category if they match each and every element of the category's definition, and fall outside the category if they mismatch on any one. For prototype theories an instance may fall within a category even if it mismatches on a number of properties.

Simple problem solving Solving problems that are small-scale and well-defined, and that require no specialist domain knowledge.

Somatic markers Physiological responses ('gut feelings') that Damasio hypothesizes may be used to guide decisions. They may be particularly influential in circumstances where it is difficult to arrive at a decision by means of conscious deliberation.

Sorting The sorting task involves participants being shown an array of different items (e.g. words printed on cards) and asked to sort them into groups. A free sort is the result of such a task where participants can sort the items into as many or as few categories as they wish.

Source estimation Source estimation involves trying to infer, on the basis of the pattern of electrical activity at the scalp surface (in the case of EEG) where the critical processing occurs inside the brain. This is sometimes known as the inverse problem.

Spatial smoothing A technique used in fMRI such that the activation at each point is replaced by a weighted average of the activation measured not just at that point, but at all neighbouring points.

SPECT Single-photon emission computed tomography or SPECT is much like PET, but with poorer spatial resolution.

Speech segmentation This is the process by which continuous speech is broken down into separate words. This is much more difficult than segmenting written language as there are not obvious breaks between words.

Spotlight of attention Items sufficiently close to one that is being attended may also receive processing, as if attention swings a spotlight and captures everything that is 'illuminated' by the imaginary beam.

Stage theories Stage theories of development, of which Piaget's is the best known, rest on the assumption that development is a discontinuous process involving distinct stages that are characterized by qualitative differences in behaviour. Stage theories can be contrasted with continuous theories, which posit that development is an incremental process.

Standardized stereotaxic coordinate system Sometimes called the Talairach coordinate system, this is a system that allows any location in the brain to be indexed by values of three coordinates.

Startle response An eye-blink reflex that protects the eye from a potential threat.

State emotion Also called mood or affect, state emotion refers to how you feel right now.

Stereopsis Due to their different positions, the retinal images of the left and right eyes will be slightly different. This difference, known as stereopsis, can be used to determine how far away an object is.

Stimulus onset asynchrony (SOA) The time between the start of one stimulus and the start of the next. Normally the first would have ended before the second commenced.

STM span (word span, digit span) The number of items (e.g. words or digits) that can be recalled after being heard or seen once; the capacity of STM.

Storage The way in which information is represented in memory.

Strength of evidence Target events or items presented and encoded in an experiment will provide a greater strength of evidence signal than other events or items in memory that were not presented and encoded in the experiment.

Stroop task The standard Stroop task involves participants naming the colour of the ink in which colour words are written. When the ink colour conflicts with the colour word (e.g. the word 'blue' written in red ink), participants are slowed down compared with stimuli where the word meaning and ink colour match (or are congruent).

Structural descriptions An approach to object recognition designed to cope with objects more complex than simple alphanumeric characters. It is based on describing objects through a set of propositions comprised of both the elements of an object and the structural relation between them.

Subjective expected utility The expected utility hypothesis is a theory of utility in which 'betting preferences' of people with regard to uncertain outcomes (gambles) are represented by a function of the payouts (whether in money or other goods), the probabilities of occurrence, risk aversion, and the different utility of the same payout to people with different assets or personal preferences.

Subliminal Stimuli presented below the participant's level of conscious awareness are said to be presented subliminally.

Subtractive designs Experimental designs often used in neuroimaging. They are based on an underlying logic that assumes tasks can be broken down into component processes. For example, if two experimental conditions, *A* and *B*, are identical except that *A* requires a particular cognitive function f while *B* does not, then one can identify those cortical regions that are involved in f by subtracting the pattern of regional cerebral blood flow (rCBF) observed during *B* from that observed during *A*.

Sub-vocal rehearsal (rote rehearsal, sub-vocalization) Rehearsal by silent speech.

Syntactic ambiguity Syntactic ambiguity occurs when the same sentence can be interpreted in more than one way. This type of ambiguity arises not from single words having multiple meanings, but from the relationship between words within the sentence. For example, the sentence 'Flying planes can be dangerous' can be interpreted as meaning that either (a) the act of flying a plane can be dangerous, or (b) planes that are flying (as compared to stationary) can be dangerous.

Syntax The principles and rules underlying sentence construction and word order within sentences are referred to as the syntactic information within the language.

Systematicity The ability to understand some sentences of a natural language brings with it the ability to understand other sentences. For example, someone who can understand 'The dog chased the cat' can also understand 'The cat chased the dog'. This is known as the systematicity of language understanding and it is explained by the fact that the sentences have compositional structure and the second sentence is built out of the same constituent parts as the first sentence. By extension, a system of representation can be described as exhibiting systematicity when the representational resources that are used to represent some states of affairs can be recombined to represent other states of affairs involving the same objects and relations.

Systems approach Theoretical approach to investigating memory that emphasizes structures or systems, which are frequently assumed to have a distinct location in the brain. This approach assumes that memory stores and memory processing are localized in the same part of the brain.

Tachistoscope Literally a 'fast-viewer', this device makes it possible to display a visual stimulus for a very brief but accurately measured period of time.

Template matching A theory of object recognition based on comparing the pattern of an object with a series of templates stored in memory of all the objects seen previously. For example, the letter 'R' would be recognized because the specific pattern of that letter would match the 'R' template in memory.

Temporal order judgement (TOJ) A task in which a participant is required to judge which of two events occurred first.

Texture gradient An example of what Gibson termed invariant information. It refers to the fact that apparent changes to the individual elements of the environment (e.g. pebbles on the beach) provide information regarding the distance, orientation, and curvature of a surface.

The Big Five Five emotions – anger, fear, sadness, disgust, and happiness – commonly considered to be basic emotions by theorists adopting the basic emotions approach.

The representational effect The phenomenon whereby problems with identical structures may be experienced as more or less easy to solve, depending upon the way in which the problem comes to be represented.

TMS Transcranial magnetic stimulation or TMS involves inducing a brief current in brain tissue through the scalp and skull, using a brief magnetic pulse delivered via a coil (about the size of a hairdryer) placed over the head. It safely and temporarily disrupts function in a specific brain region in order to evaluate the importance of that area of the brain to performance on a given task.

Top-down processing An approach to cognition that starts with existing knowledge of the environment, which is then used to guide perception.

TOTE unit An iterative problem solving strategy based on feedback loops.

Trait emotion Refers to more stable emotional or personality characteristics, i.e. 'what kind of person' you are as opposed to how you are feeling right now.

Transcranial magnetic stimulation See TMS.

Transfer The capacity to apply acquired knowledge and skills to new situations.

Transfer appropriate processing (TAP) TAP focuses on the overlap between the processes engaged at encoding and the processes engaged at retrieval. Specifically, it predicts that best memory performance will be observed when the processes engaged at encoding transfer appropriately to retrieval.

Transitivity Height provides an example of a transitive relationship. If Anne is taller than Betty, and Betty is taller than Caroline, then Anne must be taller than Caroline.

Turing test A test proposed by the mathematician Alan Turing in order to answer the question as to whether machines (computers) can be said to be intelligent. His proposal was that the responses of a machine should be compared to those of an intelligent human. If they were found to be indistinguishable – if a third party could not tell whether the responses were those of a human or a machine – then the computer should be said to be intelligent. A similar argument can be made concerning consciousness.

Type I processing Repetitive processing (as opposed to processing that proceeds through further levels), for example, verbally rehearsing a telephone number to keep it 'in mind' before calling the number.

Type II processing Processing that proceeds through further levels of increasing depth.

Type I reasoning Type 1 reasoning involves analytic processes carried out in working memory and is a late evolutionary development.

Type 2 reasoning Type 2 reasoning involves long-term memory processes involving association and heuristics.

Typicality ratings Rosch (1973) elicited participants' ratings of the typicality or 'goodness-of-exemplar' (sometimes referred to as GOE) of particular instances of a category – the method is often known as a typicality ratings method.

Uniqueness point The uniqueness point of a word is the point at which only one word in your mental lexicon matches the speech input. For example, if you hear 'dolf', only one word in your mental lexicon, dolphin, is consistent with the speech input. According to the cohort model this is the point at which the recognition process is complete.

Up-regulation Up-regulation refers to an individual's ability to increase an emotional response, and most often applies to positive emotions (e.g. making an effort to feel happy and relaxed when out partying with friends).

Valence In the context of emotion research, the positive/pleasant or negative/unpleasant qualities of a stimulus.

Variable physical realization Functionalist theories define things in terms of what they do, rather than in terms of what they are. A doorstop, for example, is defined in terms of the role that it plays, and not in terms of what it is made from or its shape or colour. Thus, the role property of being a doorstop has many different possible physical realizations, just as the role property of being Hamlet has been realized by many different actors. In cognitive psychology, states and processes are defined in terms of their role in the cognitive system and their physical realizations may vary in different individual human beings, in different creatures, and in different non-biological systems.

Ventral stream The pathway that projects from the primary visual cortex to the inferotemporal cortex. It specializes in pattern discrimination and object recognition. Sometimes referred to as the 'what system', as it deals with perception for recognition, or in other words information about what an object is.

Ventriloquism effect The tendency for speech sounds to appear to come from lips speaking those sounds, even when the actual sound source is displaced from where the face is seen speaking.

Verbal analysis A method used to gain an understanding of the representation of knowledge used in cognitive performance, and how that representation changes with learning.

Viewer-centred description A mental representation of an object that is limited to the particular viewpoint that the object was seen from.

Visual illusions Artwork or other depictions that 'fool' the senses into perceiving incorrectly. A key example is the Müller–Lyer illusion, in which two lines of equal length appear to be of differing lengths due to the addition of arrowheads pointing in opposing directions.

Visual or sensory neglect A neuropsychological condition in which the patient is only able to perceive one side of an object.

Visuo-spatial sketchpad A sub-system of working memory that functions as a buffer store for visually and spatially coded information, and possibly also kinaesthetic information.

Wavelength Waves produce repeating patterns of displacement; in the case of sound waves, small changes in the bunching together of air molecules. The distance from a point on a wave to where it next repeats itself is the wavelength. The longer the wavelength of a sound, the more slowly the eardrum vibrates and the deeper the perceived pitch.

Well-defined categories Some categories, like even numbers, seem amenable to definition. For these well-defined categories, the classical view appears to give a good explanation of category membership, though it does not obviously explain how some even numbers are considered more typical than others.

Wernicke's aphasia A language impairment mainly consisting of difficulties with verbal comprehension associated with fluent speech.

Within-category distinction One form of object recognition that is based on recognizing a specific exemplar of a category. For example, recognizing that an object is your car, rather than just recognizing that the object is a car. Perhaps the most common form of within-category distinction is face recognition, as the focus is on recognizing which specific face one is looking at.

Word-initial cohort The word-initial cohort is the set of words that are activated based on a match between the first sounds in the word and the first sounds in the stored representations of the words. For example, when the speech input 'do' is heard, the words dog, doll, and dolphin will all be activated (amongst others).

Words and rules model A dual-route model of inflectional morphology, specifically the formation of the past tense of English verbs. According to the words and rules model, formation of the past tense involves two separate components of the language system. Irregular past-tense forms are stored in the lexicon, while regular past-tense forms are generated by application of a default rule that is stored in the grammar.

Wordspotting In a wordspotting task you would be asked to detect familiar words embedded within a list of spoken non-words. For example, the word *mint* might be embedded in the longer non-word mintayve. This task can be used to investigate the processes involved in speech segmentation.

Working memory A temporary store for recently activated items of information that can be manipulated and moved in and out of short-term memory. Working memory may be thought of as a conduit to consciousness, serving to bring together information in different modalities from perception and long-term memory into conscious representations.

Working memory span (counting span, operation span, listening span, reading span) The capacity of a general working memory system, underpinning the ability to combine information processing with temporary storage in a range of different activities.

Working self A set of hypothesized higher order executive processes that are thought to operate in order to construct autobiographical memories appropriately, to keep irrelevant knowledge and memories out of mind, and to ensure that what is recalled is relevant to the currently active task or goal. The working self is hypothesized to control the construction of autobiographical memories, and is conceived as a complex hierarchy of interconnected goals, all of which are in varying states of activation, but only some of which can enter consciousness.

Yerkes–Dodson Law An empirical relationship between arousal and performance, which states that performance increases as arousal increases, but only up to a certain level. As arousal increases beyond this level, performance decreases.

Zombie A hypothetical person who is postulated to lack full consciousness but nonetheless has the behaviour of a normal human being.

This page intentionally left blank

INDEX

Page numbers in **bold** indicate the location of key concepts or terms that are either explicitly or implicitly defined in the surrounding text and/or explained in the glossary.

A
AB effect 46
abductive inference **656**
abstract conditional inference task 372–5
abstract selection task 379–82
access consciousness 579
acquisition time **453**
ACT* theory 326
ACT-R model 610
  ACT-RN 633
  architecture 611–12
  compared to PDP models 632–4
  declarative memory 612–13
  history 611
  learning and problem solving 625–32
  memory phenomena 618–25
  perception and action 616–17
  procedural memory 613–15
ACT-RN 633
acting 326
action
  ACT-R model 616–17
  attention 59
  perception for 69, 91
  rule-based systems **609**
activation **613**, 643
activation equations **622**
activation image **456**
activation level **175**
activation PET **458–9**
activation threshold **623**
active perception 104–5
addition by counting 627–9
additional antecedents 375–6, 378
addressed phonology 178
adrenalin 510
affect grid 518
affective primacy hypothesis 594–5
affirming the consequent 369
affix-stripping 183
affordance **80**
agentic personality 559
agnosia **408**
aha! phenomenon 306
air theory 74
aircraft proximity event 93–4
alcoholism 237
algorithmic level 24–5
all or nothing 376
Allais paradox **340–1**, 347
alpha waves 60
alternative antecedents 375–6, 378
Alzheimer's disease 164, 472, 474–5, 482
ambient optic array **74–9**
amnesia
  anterograde **401**
  childhood 550
  developmental (early onset) **241**
  implicit memory 585
  PTSD 567
  retrograde **401**
amnestic syndrome **236–7**, 240–1, 242
amygdala 496, **516**, 540–1
anaesthesia
  judging depth of 597–8
  learning during 588–9
analogical mapping **318**
analogical paradox 317
analogical problem solving 316–19
anaphora 198, 199–200
anatomical hypothesis **456**
anatomical landmarks **455**
anatomical syndromes 410–13
anchor and adjust heuristic **354**
anergia 486
angiograms **452**
anhedonia 486
annihilation reaction **458**
anosognosia **422**
antecedent 368
anterior commissure (AC) **454**, 455
anterograde amnesia **401**
anti-aircraft gunnery 14
Anton's syndrome 472
anxiety-related attentional bias 530–2
appraisal theories 538–9
apraxia **401**
architectural assumptions **624–5**
Aristotle 366
arousal **518**
artefacts **437**, **440**
articulation 217
articulatory loop **272**
articulatory suppression **275–6**
artificial grammar learning 587–8
artificial intelligence 16, 581
assembled phonology 178
associated words **185**
association 414
association strength 185, **622**
at that point **200**
Atkinson and Shiffrin model 231, 268
attention 33–64
  auditory 33–40
  consciousness 60–1
  cross-modal 54–5
  directing actions 59
  distraction 53–5
  divided 282
  emotion 530–2
  executive control 281–2
  focused 282
  integration, **44**, 48–53
  masking 47–8
  neurology 55–8
  spotlight of **36**
  switching 282
  visual 40–8
attentional blink **46**
attenuation **38**
attribute-listing **147**
attribution theory **534**
attributional bias 481
audience design **211–13**, 214
auditory attention 33–40
auditory hallucinations 479–80
auditory oddball paradigm **444–6**
autism 164
autobiographical knowledge 554–6
autobiographical memories 546–52
  constructing 560–3
  distress 564–9
  recollection 558
  self 570–1
automatic processing 578, 589–91, 593
autonomic nervous system **509–10**
autonomous view 186
autonomy 673–4
autoscopic hallucinations 480
auxiliary assumptions **624**
availability heuristic **354**
average reference **436**
averaged **437**
avoidance **567**
awareness 255
away sentences **203–4**

B
backward masking **42**, 44, 94–5
backward propagation of error **608**
backward recall **618**
backwards search **310**
Balint's syndrome 57
band-pass filter **440**
bar 84
base-level activation **622**
base rate neglect **353**, 354–5
baseline correction **440**
basic emotions **512–15**
battery of tests 424
Bayes' theorem 349–51
Behavioural Assessment of the Dysexecutive Syndrome (BADS) scale 471
behavioural decision research 342–5
behavioural response measurement 423–4
behaviourism 10–12, 581
belief formation 656–7
between-category distinctions **102**
biases 352–6
bi-conditionals 373
Biederman's theory 114–17
Big Five emotions **513**
binding of item and context (BIC) model 251–2
binding problem **283**, **600**
bits of information 381
blind spot 90
blindsight 409, 592
blob 84
block design **515**

blood oxygen level dependent contrast **451**, 452
blood oxygenation **447**
BOLD contrast **451**, 452
book bag paradigm **349–51**, 352
borderline cases **149**
bottom-up processing 9, **70**, 73, 82, 95
bottom-up support **175**
bounded rationality **360**
box and arrow diagrams 405–6
box and candle problem 308, 313
brain
  anatomy and functional areas 404
  focal and diffuse injury 470–3
  functional reorganization 409–10
  traumatic injury 488–90
brain reading 458
brain-stem injury 472–3
brain-stem responses **437**
Brentano, Franz 20–1
bridging inference **201**
Broadbent, Donald 37
Broca, Paul 404–5
Broca's aphasia **406**
Bruce and Young face recognition model 123
buffers **612**

C

calibration tasks 356–7
candle and box problem 308, 313
Cannon-Bard theory 536–7
canonical coordinate frame **109**
Capgras delusion/syndrome **65**, 128, **482**, 483, 522–3
capture error 281, 616
car parking problem 315–16
Cartesian theatre **591**, 601
categories **142**
categorization 101, **142–4**, 145, 146–61
  classical view **146–51**
  commonsense theories **155–8**
  development 158
  by different people 164
  different ways 162–3
  exemplar theories 155
  expert opinion 160–1
  judgements **149**
  language 163
  modes of 162–3
  prototype theories **151–5**
  psychological essentialism 158–61
  similarity 146–55, 156, 163
category-specific impairment **3**, 4–5, 6
category verification **147**
cause-effect explanations 13
central executive 269, **272**, 279–85, 600
central masking 44
central systems 656–9, 664
central workspace 272, 280–1
cerebellar affective syndrome 473
cerebellar injury 472–3
cerebral cortex **474**
chaining hypothesis 288
characteristic-to-defining shift 158
Cheshire cat illusion 599
chess playing 24, 274, 320–1, 323–4, 325, 329
childhood amnesia 550
Chinese Room **597**
Chomsky, Noam 12–13
chunks 274, 284, **612–13**, 619–20, 625
classical dissociation 416–17
classical view of concepts **146–51**
cleft sentence **205**
closure 71
cluster analysis 4
clustering 85
coarticulation **171**
cocktail party effect **34**
co-evolution of theories **670–2**
cognitive appraisals **537**, 540
cognitive approach 20–3
cognitive architectures 610
cognitive behaviourists 13
cognitive bias **525–6**
Cognitive Bias Modification (CBM) **525–6**
Cognitive Bias Modification for Attention (CBM-A) 532
cognitive correlates of consciousness **578**
cognitive modelling 122–4, **606–11**
cognitive modules 385, 388
cognitive neuropsychiatry 468–502
cognitive neuropsychology **399–429,** 663–8
  assumptions 407–14, 663
  compensatory strategies 420–1
  historical milestones 404–7
  inference 664–6
  lesion method 399–404, 432
  methods 414–20
  past-tense debate 666–8
cognitive psychology
  characteristics 2–8
  history of 8–17
  sub-disciplines 3
  subject matter 5
cognitive reframing 485
cognitive style 161
cognitive system **122**
coherence **198–9**, 283
cohesion 198–9
cohort model 173–4, 175
coin manipulation problem 314–15
collinearity 115
common ground **211–13**
commonsense theories **155–8**
communal personality 559
comparability **339**
compensatory strategies 420–1
competition approach 175–6
complex concepts **154–5**
component axes 111
components **437**
components of processing (COP) account 256–7
compositional structure **649–51**
computation 21–3, 641–51
computational level 23–4
computational model 6, 287–8
computational model of the mind (CMM) **642–3**
computer analogy 581
computers 15–16, 18
COMT enzyme 479
concepts **141–2**
  animals 143
  classical view **146–51**
  cognition 144–6
  complex concepts **154–5**
  different types of 163–4
  explaining categorization 143
  inferences 145
  prototype theories **151–5**
  reasoning 145
  remembering information 145–6
  semantic memory 145
  theory-theory 155, 157–8
  words and 142, 145
conceptual incidental tests **253**
conclusion 367–8
condition **609**
conditional 368
conditional inference 372–9
conditional probability 371
cones 68, 90
confabulation 402–3
confidence intervals 358–9
configural processing **132–3**
confirmation bias 380
conjunction fallacy **353–4**, 355–6
connectionist modelling 22–3, 643–4
  category-specific impairment 6
  face recognition 124–5
  mental representation 649–51
  past tense 645–7
  rules 647–9
connectionist systems **236**
connections 643
connectives 368
consciousness 9, 577–605
  access 579
  altered states of 597–9
  attention 60–1
  cognitive correlates **578**
  cognitive psychology 581–2
  cognitive studies 583–92
  cognitive theories 600–2
  defining 578–80
  detecting 597
  easy and hard problems 579
  explanatory gap 579, 580
  functions 593–9
  global workspace theory **600**, 601–2
  higher order thought (HOT) theory **600**, 602
  modularity 591–2, 596
  multiple drafts theory **601**
  neuropsychology 591–2
  phenomenal 579
  philosophical approaches 580–1
  stream of 579
  working memory 600–1
consensus 215–16
consequent 368
conservatism **351–2**
constraint relaxation **313**
constructability **245**
constructivist approach **87–90**
content words **181**
context and typicality 153–4
continuous EEG 433
contour generator **110–11**
contrastive analysis 577
controlled processing 578, 589–91, 593
conversion disorder 496–7
correlational designs **449**
Corsi span 273
cotermination 115
counting span 280
covert recognition **127**
criterion for satisfactory progress theory 313–16
cross-linguistic differences, memory span 277
cross-modal attention 54–5
cross-modal priming **174**
curvilinear aggregation 85
curvilinearity 115
cut-off frequency **440**
cybernetics 14
cyclotron **459**

D

data analysis **456**
decision analysis 335, **336–8**
decision making
  behavioural decision research 342–5
  descriptive theories **335**
  experience 348–9
  fast and frugal **360–1**

  - judgement and 334–5
  - normative theory **335**, 336–42
  - prospect theory **346–9**
  - theories 335
- decision tree **336–8**
- decision weight 347
- declarative knowledge **237**
- declarative memory **611**, 612–13
- declarative memory system **237**, 238–43
- decoding 458
- decompositional approach 183
- deductive reasoning 368–70, 387
- deductive validity 387
- deep dyslexia 411–12
- deep structure 322
- default network 457
- defeasible inferences 375
- definitions **146**, 150–1
- delusions 481–4
- dementia 473–7
- dementia with Lewy bodies (DLB) 475–6
- denying the antecedent 369
- deontic conditional 383
- deontic selection task 382–6
- dependency goal chunks **625–6**
- dependent measure **447**
- depolarization **82**
- depression 481, 484–90
- depressive realism 534
- depth of processing **231**
- depth perception **74**
- derivational morphology **183**
- descending pathways 671
- descriptive theories **335**
- detachment 565–6
- development
  - categorization 158
  - memory span 276–7
  - speech segmentation 173
- developmental amnesia **241**
- diagnosis 145, 305
- diagnostic value **245–6**
- diagnosticity **351**
- diagram-makers 405–6
- dialogue 214–18
- dichoptic **44**
- dichotic listening **37**, 38 39
- dimensional view of emotions **512**, 518–20
- dipole models **442**
- dipoles **436**
- direct perception **73**
- direct retrieval **562**, 563
- discourse 143, 198–207
- discourse-as-process 199
- discourse-as-product 199
- discriminability **245**
- disinhibitory psychopathologies 491–5
- dissociation 414–17, 565–6
- distinctiveness **232**
- distraction 53–5
- distractors 49
- distributed source models **442**
- divided attention 282
- domain knowledge 326
- domain specificity 407, **653–4**, 665
- dominance (emotions) 518
- dominance (hemispheric) 471
- dominance (options) **339**
- dopaminergic system 479, 481, 486, 493
- doppelganger 480
- dorsal stream 57, **91–3**
- dorsal visual-limbic pathway 128
- dot probe **530–1**
- double dissociation 415–17, **663**, 655–6
- down-regulation **541**
- DRC model 178–9, 642
- driving and mobile phone use 55
- drug-induced perceptual disturbances 598
- dual process signal detection theory (DPSD) 248–50, 258
- dual process theories of reasoning 389
- dual-route models reading 178–9, 414
- dual-task paradigm **269**, 273
- dyslexia **411–12**

E

- E-FIT images 102
- early onset amnesia **241**
- echoic memory **37**
- ecological optics **74**
- edge-segment 84
- 8-coin problem 314–15
- elaboration 233, **313**
- elaborative inferences 201–2
- electrocardiogram interpretation 305
- electrodes **435**
- electroencephalogram (EEG) **45**, 430, **433–46**, 465
  - basic data analysis 436–40
  - recording 435
  - resolution 431, 432
  - statistical data analysis 443–4
- electromyography 510
- electrooculogram **440**
- embodied cognition **203**
- embodiment 202–4
- emergent property **609**
- emotion 507–45
  - attention 530–2
  - basic emotions **512–15**
  - behaviour 508–9
  - Big Five **513**
  - bodily responses 509–10
  - communication 521–2
  - components 508–12
  - depression 484–5
  - dimensional view **512**, 518–20
  - feelings 511, 523–4
  - function 520–4
  - goals 520–1
  - hot and cold components 525
  - imaging 515–17
  - influenced by cognition 535–42
  - influencing cognition 524–35
  - information 522–3
  - manifestation of 525
  - memory 526–9
  - modal emotions 518
  - processing of 525
  - semantic interpretation **532–4**
  - state emotion **524**
  - trait emotion **524–5**
  - verbal labels for 517–18
- emotion family **512**
- emotion regulation **541–2**
- emotional facial expressions 102, 104, 126–7
- emotional Stroop task 51, 530
- encoding **230–5**
- encoding specificity **243–5**
- endowment effect 347
- entorhinal cortex 241
- EOG **440**
- epiphenomena **581**
- episodic buffer **284–5**
- episodic information **43**
- episodic memory **229–30**, 238, 240–3, 556–8
- episodic record 277, 283
- errorless learning 595–6
- essentialism 158–61
- evaluability **345**
- evaluability principle **345**
- event-related designs **516**
- event-related field **437**
- event-related potentials (ERPs) **58**, 185, 186, **437**
- evoked magnetic fields **437**
- evoked potentials **437**
- evoked signals **438**
- evolutionary psychology 382, 385–6, 388–9
- exchange errors 208, 209–10
- executive functions 279–85, 470
- exemplar theories 155
- expected information gain 381–2
- expected utility 336, 337–8, 384, 385
- experience-based decisions 348–9
- experimental design 447–9
- expert-novice paradigm 324
- experts/expertise
  - categorization 160–1
  - clinicians **305**
  - face recognition 130–3
  - homogeneity 328–9
  - modal model 324–6
  - overconfidence 356–7, 358
  - problem solving 320–4
- explicit memory **252**
- extended 10/20 system **435**
- external representations **309–10**
- eye 68
- eye movement artefacts **440**
- eye movements when reading 180–2
- eye muscle artefacts **440**
- eye-tracking techniques 180–1

F

- face preference of newborns 130
- face recognition **102–4**
  - Bruce and Young model 123
  - cognitive model 122–4
  - compared to object recognition 118–19
  - connectionist model 124–5
  - covert recognition **127**
  - everyday errors 121–2
  - familiar and unfamiliar faces 102, 119–21
  - IAC model **124–5**, 127, 129
  - inversion effect **130–3**
  - modelling 121–6
  - specialness of faces 129–33
  - Young's diary study 121–2
- face recognition units **122**, 124–5
- facial emotion recognition 102, 104, 126–7
- factor analysis 479
- factorial design **448–9**
- false fame effect 584–5
- false memory 402–3
- familiarity 246, 249, 250
- family resemblances 150
- fan effect **622–3**
- fast and frugal decision making **360–1**
- featural theory **184**
- feature binding 284–5
- feature integration theory **50**
- feature recognition theories **106**
- feedback **13–15**, 60–1, **104**
- feelings 511, 523–4
- field dependence 161
- field inhomogeneities **452**
- figure-8 coil **462**
- filter **38**
- filtering **440**
- fired **609**
- first-order relational properties 131
- fixations **180**, 181
- 'flanker' effect 50–3
- flashbulb memories 547
- flow in optic array 78–9
- flow of information 69–70
- fluent dysphasia 472
- fluorodopa **460**

focused attention 282
Fodor, Jerry 407, 652; *see also* modules and modularity
form 370
formal rule theories 371
formal thought disorder 480
format dependence 358
forward recall **618**
forwards search **310**
fovea **68**
fractionation 282–3, 401, 405, 408, 681
frame problem **657–9**
framing effects **348**
free recall **232**, 233, 234, **618**
frequency **355–6**
frequency power spectrum **438**
frontal lobe injury 470–1
frontal patients 281
Frontal Systems Behaviour Scale 471
frontal variant fronto-temporal dementia 476
fronto-temporal dementia 476
full-listing approach 183
full primal sketch **85**
function words **181**
functional accounts **25**
functional connectivity analysis 456–7
functional dissociations 255
functional fixity **308–9**
functional magnetic resonance imaging (fMRI) 430, **447**, 449–57, 465
  advantages 453
  data analysis 456–7
  emotion processing 515–17
  face recognition 130
  image processing and analysis 453–6
  limitations 452–3
  living/non-living distinction 4
  mechanism 450–2
  resolution 431–2
  speakers and listeners 220–1
functional syndromes 410–13
functionalist psychology **10**, 581
fusiform gyrus 130
fuzzy categories **164**

G

Gage, Phineas 404
galvanic skin response (GSR) **39**, 510
gambles **336**
gamma camera **460**
gamma oscillations **438**
gamma rays **458**
gamma synchronization 60
garden path model 190
garden path sentences 269
Gaussian blurring **83**
general events **555**
general theory of the mind 641–2
generalization **622**
generalized cones **109**
generalizing 23
generative retrieval **561–2**, 563
geons **114–15**
Gerstmann's syndrome **413**
Gestalt movement 9–10, **71–2**, **306**
Gibson's theory of perception 73–81
global workspace theory **600**, 601–2
goal attainment **555**
goal buffer **615**
goal-directed explanations 13
goal hierarchy **548**
goal module **612**, **616**
goals 615–16
  emotions 520–1
  learning and problem solving 327–8
  working self 558–60
good continuation 71
good models 634–6
goodness of exemplar 147
graceful degradation **609**
gravitational force 10, 11
Gregory, Richard 87–90
grey level description 82
greyscale 82
ground theory 74
group **619**
group decision making 218–19
group studies 417–19

H

habituation **516**
haemodynamic lag **452**
haemodynamic response **451**
hallucinations 60, 61, 479–80
hands, exploratory movements 104–5
haptic information **104–5**
harm avoidance 494
hazard management 388–9
head-up display (HUD) **41**
hearing tests **438**
hedonic homeostatic dysregulation syndrome 486
Helmholtz principle **440**
hemispheric dominance 471
heuristics **311**, 352–6
hidden layer **607**
hierarchical subtractive design **448**
higher order thought (HOT) theory **600**, 602
hill-climbing **313**
Himba 116
hippocampus 236, 237, 241, 242, 572
history
  cognitive neuropsychology 404–7
  cognitive psychology 8–17
Hixon symposium 12
HM 236–7, 401
hoarding 493
hobbits and orcs problem 309, 311, 312
hollow faces 89
homographs **533**
homonyms 185–6
homophones **179**, **533**
homunculus problem 601
horizon ratio relation **74–5**
hotspots 569
Humphrey and Bruce model 101
hybrid models **611**
hyperarousal 567
hyperimportance 481
hyperopia of the future 494
hypersalience 481
hypnosis 59, 60, 61, 598
hypothesis confirmation 656–7

I

IAC model **124–5**, 127, 129, 177
IAM task 570–1
iconic memory **41**
if p then q 368
'if. . .then' 366–7
image time series **453**
imaging studies, *see* neuroimaging
impasse **313**
implementational level 25
implicit cognition 583–9
implicit learning **173**, 586–9
implicit memory **252–7**, 583–6
implicit negation 380
implicit processing 412–13
impoverished figures 88–9
impulse control disorders 493–5
incidental tests 252–3
incremental interpretation **200**
independence **339**, 340–2, 673
indicative conditional 383
indirect memory test 583–4, 585
indirect perception 73
indirect speech acts 200–1
individual differences
  executive function 282–3
  problem solving 326–9
  reasoning 389
  vocabulary acquisition 286
induced signals **438–9**
inductive strength 387
inertia effect **351**
inference
  cognitive neuropsychology 664–6
  conditional 372–9
  reasoning 145, 367
  text interpretation 201–2
inferotemporal cortex 4
inflectional morphology **183**
information gain 381
information processing
  insight 311–13
  problem solving 310–11
information transfer 216, 218
informational encapsulation 408, **653**, 654–5, 659–61
inhibition 281
inion **435**
initiate **204**
inner-loop monitoring 213–14
input layer **607**
insight **306**, 311–16
integration **44**, 48–53, 233
intentional tests 252–3
intentionality
  implicit memory 254–5
  mind 681
interactive alignment 216–17, 218
interactive view 186
interhemispheric fissure **455**
internal representations **309–10**
internal structure **148**
interpretive bias 533
interruption **44**
interval estimates 358
intransitivity **149–50**, 339–40
introspection **9**, 581, 582
intrusions 566, 568–9
intuition 585
invariance **339**
invariants **74–8**
inverse problem **440–3**
inversion effect **130–3**
Iowa gambling task **523**
irreducibility 673
irregular word reading 178–9
irrelevant speech 53–4, 277
isolated forearm technique 597–8
isomorphic **309**
isosensitivity curve 247
isotropic **656–7**
item specific processing 233–4

J

James, William 10
James-Lange theory 535
judgement
  Bayes' theorem 351–2
  decisions and 334–5
  heuristics and biases 352–6
  overconfident responses **356–9**
  uncertainty 349–59
jumping to conclusions 481

K

Kanizsa's illusory square 65, 66, 67
kinesthesis **104**
know experience **250–2**

knowledge
  expert problem solving 320–4
  perceptual hypotheses 88–90
  relating language to 202
Korsakoff's syndrome 237, 473

L

language
  behaviourist explanation 12
  categorization 163
  connectionist models 22
  knowledge 202
  rule-based learning 12–13
  symbolic models 22
language of thought hypothesis 650
Lashley, Karl 12
late responses **438**
lateral geniculate nucleus (LGN) 90
law of closure 71
Law of Pragnanz **72**
law of proximity 71–2
law of similarity 71–2
Lazarus, Richard 540
learning
  ACT-R model 625–32
  behaviourist explanations 11
  good and poor learners 327
  implicit **173**, 586–9
lesion method 399–404, 431, **540**
levels of explanation 23–6, 669–70
levels of processing (LOP) 231–2, 255, 258
Lewy bodies 476
lexical competition 175–6
lexical concepts **145**
lexical decision task 4, **38–9**
lexical effects **177**
lexical output 125
lexical segmentation models 170, 172–3
Lichtheim's model 406
lie detectors 39
life scripts 551
life story **556**
lifespan retrieval curve **548–52**
lifetime periods **555–6**
likelihood ratio **350**
Linda problem 353–4, 355–6
lip-reading 54–5
list memory **618–24**
listening span 270
literally **201**
living/non-living distinction 3, 4–5, 6
logic 366, 368–70
logically valid inference 369
long-term memory **229**
  compared to short-term memory 267–9
  encoding **230–5**
  retrieval **230**, 243–52
  storage **230**, 235–43
  working memory and 274
long-term working memory 325–6
loss aversion **346–7**, 348
loss of activation error 616
low-pass filters **440**

M

macula lutea 68
magnetic force 10–11
magnetic resonance imaging (MRI) 449–50
magnetoencephalogram (MEG) 430, **433–46**, 465
  basic data analysis 436–40
  recording 435–6
  resolution 431
  statistical data analysis 443–4
magnocellular pathway 90
mandatory modules 407
mapmakers 405–6
Marr, David 23–6, 81–7, 669
Marr and Nishihara's theory 109–14
masking
  attention 47–8
  backward **42**, 44, 94–5
  central 44
  peripheral 44
massive modularity hypothesis 661–2
matching bias 380
matching effect 380, 381, 382
matchstick algebra 313
materialists 580
maximization heuristic 314
may 384
meaning 370
means-ends **310**, 311, 327
medical diagnosis 145, 305
memory
  ACT-R model 618–25
  chess skill 325
  declarative **611**, 612–13
  declarative memory system **237**, 238–43
  echoic **37**
  emotions 526–9
  episodic **229–30**, 238, 240–3, 556–8
  explicit **252**
  false 402–3
  flashbulb 547
  iconic **41**
  implicit **252–7**, 583–6
  intrusions 566, 568–9
  lists **618–24**
  modal model 268
  mood congruent (MCM) **526–8**
  mood dependent **528–9**
  multiple memory systems 238–43, 256
  multi-store model **231**, 268
  non-declarative memory systems **237**, 238–43
  personality 559
  procedural **611**, 612, 613–15
  processing approach 230
  recognition **232**, 233, 234, 246–50
  self 571
  self-defining 547, **560**
  self-memory bias 235
  semantic 145, **238**, 240–3, 556–8
  short-term **231**, 267–9
  systems approach 230
  see also autobiographical memories; long-term memory; working memory
memory span 267, 276–7
memory stores 235–43
memory systems 235–43, 256
mental acts 20–1
mental arithmetic 267
mental contents 20–1
mental heuristics **352**
mental lexicon 145, **170**, 182–7
mental logic 370, 371, 373, 376–7, 380, 384, 387
mental models 202, 370, 371, 373–4, 377–8, 380–1, 384, 387–8
mental representation 640, 649–51
mental states 20–1
mentalistic terms 580
mere exposure **539**, 586
metaphor 318–19
metrical foot **171**
middle latency responses **438**
mild cognitive impairment 476–7
mind-body problem **580**
mind of the reader **199**
mindfulness-based cognitive therapy 527
mirror agnosia 483
mirrored-self delusion 483
mismatch 45
mismatch negativity (MMN) **444**
missionaries and cannibals problem 309
mobile phones
  texting skills 629–31
  using when driving 55
modal emotions 518
modal model
  expertise 324–6
  memory 268
model development 17–18
modules and modularity **82**, 407–8, **640**, 651–6, 665, 681
  central systems 656–9, 664
  consciousness 591–2, 596
  features of modules 407–8, 652–5
  massive modularity hypothesis 661–2
modus ponens 369
modus tollens 369
monism 580
monothematic delusions 482–4
'monster' problem 309
mood 486
mood congruent memory (MCM) **526–8**
mood dependent memory **528–9**
morphemes **182**
morphology **182–4**
Moses illusion 204, 205
motion parallax **78**
moving window experiment 181–2
Müller-Lyer illusion 65, 66, 67, 408
multinomial processing tree model 390
multiple comparisons correction **456**
multiple drafts theory **601**
multiple memory systems 238–43, 256
multi-store model **231**, 268
must 384
myopia for the future 494

N

N400 186
naïve sampling model 359
naming 101, 102
nasion **435**
near-infrared optical spectroscopy (NIRS) 430–1, **447**, 460–1, 465
necessary inference 201
Necker cube 65, 66, 67
negative symptoms 479, 480–1
neglect syndrome 56–7, 65, **410–11**, 412–13
neighbouring words **179**
neurobiology 669, 672, 673–4
neurochemicals **457**
neuroimaging 18–19, **236**, 430–67
neuron doctrine 672
neuropsychological data **236**
neuropsychology 400–1, 663
neurotransmitters **460**
Newell Test **635**
9-dot problem 307–8, 313–14
nodes **607**
noise 439–40
non-accidental properties 115–16
non-associative models 289
non-declarative memory systems **237**, 238–43
non-deductive reasoning 387
non-demonstrative inference **656**

- non-literal meaning 200–1
- non-target effects 50
- normality 3
- normative theories **335**, 336–42
- novelty seeking 494

O

- object-centred description **85**, 117
- object recognition **102–4**
  - Biederman's theory 114–17
  - Humphrey and Bruce model 101
  - Marr and Nishihara's theory 109–14
  - three-dimensional objects 108–18
  - by touch 104–5
  - two-dimensional objects 105–7
  - viewpoint 107–8
- obligation rule 383
- observation 10
- obsessive compulsive disorder 491–3
- occipital lobe injury 472
- occluding contours **109**
- occlusion 78
- Ockham's Razor **635**
- oddball paradigm **444–6**
- operation span 271, 280
- operationism **11**
- operators **311**
- optic array 74–9
- optimal viewing position (OVP) **181**
- organization 12–13
- orthography **178**
- outer-loop monitoring 213–14
- output layer **607**
- overconfident responses **356–9**
- own name effect **38**, 47

P

- P300 component **444**
- paediatric autoimmune neuropsychiatric disorder (PANDAS) 491, 492
- Pandemonium system 106
- parafoveal region **181**
- parahippocampal cortex 241
- parallel activation **173–4**
- parallel distributed processing (PDP) **607–9**, 632–4
- parallel processing **38**, 39, 43
- parallel search 48–50
- parallelism 115
- parameter estimates **456**
- parametric designs **449**
- parasympathetic ANS **511**
- parietal lobe injury 471–2
- parietal region **49**, 56, 57, 278, 279
- parkinsonian symptoms 476
- Parkinson's disease 473, 481, 487–8, 494–5
- parsing **188**, 189–93
- partial information **246**
- partial matching **623**
- partial report superiority effect **41**
- partonomic knowledge **556**
- parvocellular pathway 90
- past participle **192**
- past-tense debate 631, 632–3, 644–7, 666–8
- pattern associator **645**
- pattern classification 457
- pattern recognition hypothesis 324
- Penrose triangle 89
- perception 10, **67**
  - ACT-R model 616–17
  - approaches to 69–70
  - for action 69, 91
  - for recognition 69, 91
  - meanings 67
  - sensing 67
  - *see also* visual perception
- perceptual classification 101, **144**
- perceptual hypotheses **87–90**
- perceptual incidental tests 253
- perceptual organization **71–2**
- perceptual span 181–2
- peripheral masking 44
- permission rule 383
- person identity nodes (PINs) **122**, 125
- personality
  - frontal lobe damage 470
  - memory 559
- perspective in communication **204**, 206–7
- phantom limb 409
- phenomenal consciousness 579
- phonemes **171**
- phonemic similarity 267, 275
- phonetic feature **175**
- phonological loop 272, 274–9
  - modelling 287–90
  - vocabulary acquisition 285–7
- phonological representation **172**
- phonological working memory 274–9
- phonology **178**
- phrase structure grammar **189**
- phrase structure rules **189**
- pilot's headphones 37
- place-holder 159
- place tokens 85
- planning 12–13
- platypus 144
- pneumoencephalogram **430**
- poker chip paradigm **349–51**, 352
- polyglots 286
- polysemous words 185
- positional confusions **623**
- positional hypothesis 288–9
- positive symptoms 479
- positron emission tomography (PET) 430, **447**, 458–60, 465
  - emotion processing 515–17
  - living/non-living distinction 4
  - phonological loop 279
  - resolution 431–2
- positrons **458**
- post-traumatic stress disorder (PTSD) 547, 564–9
- posterior commissure (AC) **454**
- posterior odds **350**
- power law of practice 323
- practice 323, **615**
- pre-auricular points **435**
- prediction 220–1
- preference reversal phenomenon **343**
- prefrontal cortex 91, 242
- pre-lexical segmentation models 170–2
- premises 367–8
- 'pre-potent' responses 282
- prescriptions 383
- prescriptive approach 335, 336–8
- primacy debate **539**
- primacy effect **415**, **618**
- primacy model 289
- primal sketch 82–5
- primary task **275**
- primary visual cortex 90
- prime **174**
- priming **38–9**, **585–6**
  - affective 594–5
  - cross-modal priming **174**
  - morphemes 183
  - repetition 117
  - semantic 185, 481
- primitives **84**, **111**
- principle of pragmatic modulation 377
- prior odds **350**
- probabilistic approach 371, 374–5, 378, 381–2, 384–5, 388
- probability 349–51
- problem-reduction **310**
- problem solving 301–33
  - ACT-R model 625–32
  - analogies 316–19
  - complex problems 319–24
  - defining a problem 303–4
  - experts 320–4
  - individual differences 326–9
  - protocol analysis 304–5
  - search process **310–11**
  - simple problems 306–16
- problem space 310
- procedural knowledge **237**
- procedural memory **611**, 612, 613–15
- process accounts **25**
- processing approach 230
- production compilation **612**, **614–15**, 626–7
- production rules **609**, 613–14, **620–2**
- productivity **187**
- progress monitoring theory 313–16
- progressive deepening 311
- progressive non-fluent aphasia 476
- prominence effect **344**
- property-listing **143–4**, 147
- propositions 368
- proprioception **104**
- prosopagnosia **65**, **126**, 127, 130
- prospect theory **346–9**
- protocol analysis 304–5
- prototype theories **151–5**
- pseudodementia 490
- pseudodepressive 470
- pseudohomophones **178**, 179
- pseudopsychopathic change 470
- psychological essentialism 159–61
- psychotic illness 477–84
- purposive explanations 13

Q

- qualia **579**
- Quinean **656–7**

R

- raclopride **460**
- radial dipoles **436**
- radical neuron doctrine 672
- random generation 282, 593
- rapid serial visual presentation (RSVP) **44–7**
- rational analysis theory **611**
- rationality 367, 389
- raw primal sketch 82–5
- reaction time (RT) **530**
- reading
  - dual-route models 178–9, 414
  - eye movements 180–2
- reading span 270, 271, 280
- realigned **454**
- reason-based theory of choice **345**
- reasoning 366–92
  - concepts 145
  - deductive 368–70, 387
  - dual process theories 389
  - everyday life 367–8, 375–8
  - individual differences 389
  - integrative approaches 389–90
  - non-deductive 387
  - psychological theories 370–2
  - suppression effect 375–8
- recall-reconstruction paradigm 320, 321
- receiver operating characteristic (ROC) curve **247**

recency 552
recency effect **415**, **618**
recognition
  basics of 100–1
  perception for 69, 91
  types of 102–8
recognition heuristic **361**
recognition memory **232**, 233, 234, 246–50
recollection **248**, 249, 558
reconstructed **453**
recording reference **435**
reduced relative clause **191**
reductio ab adsurdum 373
reduction 670–2
re-encoding **313**
re-entrant pathways **95**
re-entrant processing 60–1, **95**
re-experiencing 566
reference (cohesion) 198
reference (EEG recordings) **435**
reference atlases **455**
referential communication 211, 212
refractory period **46**
regressive saccades (regressions) 181
regular word reading 178–9
relational processing 233–4
relative distinctiveness **245**
remember experience **250–2**
reminiscence bump 550–2
repairs **213**
repetition priming 117
repetitive transcranial magnetic stimulation 93
representational change theory 313, 315–16
representational similarity 457
representations **20–1**, 306–10, 643
representativeness heuristic **353–5**
requisite decision modeling **338**
resolution **431–2**
resonance 80
response bias **533**
resting metabolism **458**
restructuring 306, 311, 313
retina **68**
retinal ganglion cells 90
retrieval **230**, 243–52
retrieval intentionality criterion **254–5**, 259
retrieval mode **558**
retrograde amnesia **401**
rhyme judgements 180
rhythm of speech 171–2
risk communication 206, 207
risky choice 336–42
rods 68, 90
routine 217
routinization **217–18**
rule-based learning, 12–13
rule-based systems **609–10**
rules 640, 647–9

S
saccades **180**, 181
sampling interval **435**
sampling rate **435**
satisficing **360**
scale compatibility hypothesis 343–4
Schachter-Singer theory 537–8
schema **556**
schizophrenia 60, 472, 477, 478–9, 480, 481–2
schizotypy 164
science 10–11, 17–19
scientific confirmation 656–7
search process **310–11**
second-language learning 286
second-order relational properties 131–2
secondary task **275**
self 571
self-defining **551**
self-defining memories 547, **560**
self-defining moments **560**
self-memory bias 235
self-monitoring 213–14
self-serving attribution bias **534**
semantic ambiguity 185–7
semantic categorization task 4
semantic classification 101, 144
semantic dementia 476
semantic features **184**
semantic information **43**
semantic information units 125
semantic interpretation **532–4**
semantic memory 145, **238**, 240–3, 556–8
semantic network **528–9**
semantic priming 185, 481
semantic representations 184–5
sensation **67**
sensory neglect **56–7**, **65**
sentence comprehension 187–93
sentence superiority effect 284–5
sentence verification **147**
serial order 288–90
serial processing **38**
serial search 48–50
serotonergic system 494
servomechanisms 13–14
set effects **307–8**
shadowing (dichotic listening) **38**
shadowing (head) 35
shallow processing **204–6**
short-term memory **231**, 267–9
signal detection theory **246**
similarity 146–55, 156, 163
single-case studies 406, 419–20
single photon emission computerized tomography (SPECT) **447**, 460, 465
single process signal detection (SPDS) 246–8, 258
skin conductance response (SCR) **127**, 510
slot **612**, 620
slow responses **438**
social contract modules 385
somatic markers **523**
somatoparaphrenic delusion 482–3
sorting task **143**
soul searching 338
sound localization 34–6
spatial normalization **454**
spatial resolution **431–2**
spatially smoothed **455**
speech, irrelevant 53–4, 277
speech errors 208–10, 213
speech segmentation **170–3**
spelling – sound mapping 178–80
spin in communication 206–7
split-brain 421
spoken word recognition 170–6, 179–80
spontaneous EEG 433
spotlight of attention **36**
spreading activation models 184
SQUIDs **436**
standard space **454**
standardized stereotaxic coordinate system **454**
startle response **510**
state emotion **524**
state-space diagrams 309
statistical thresholding **456**
stereopsis **85**
stimulus onset asynchrony (SOA) **42**, 44
storage **230**, 235–43
stranding 208–9, 210
stream of consciousness 579
strength of evidence 246–7
stroke 496–7
strong artificial intelligence 581
strong dissociation 416–17
Stroop task 51, 407–8, **530**
structural alignment **318**
structural codes 122
structural descriptions **106–7**
structural MRI 450
structure mapping engine 318
structure-mapping theory 318
sub-cortical structures **474**
subjective expected utility (SEU) **336**, 339
subliminally **539**, 586
subordination 205
sub-symbolic models **633**
subtractive designs 447–9
subtractivity 408–10, 663
sub-vocalization **276**
superconducting quantum interference devices (SQUIDs) **436**
supervisory attentional systems (SAS) 281, 600–1
suppression effect 375–8
suppression training 527
supraliminal presentation 587
supramarginal gyrus 279
surface dysgraphia **414**
surface dyslexia **414**
surface structure 322
Sylvian fissure **455**
symbol grounding problem **202–3**
symbol systems 21–2
symbolic representations **633**
symmetry 115
sympathetic ANS **510**
synapse **541**
syndromes 410–13
syntactic ambiguity **185**
syntax 188–9
systematicity **649–51**
systems approach 230

T
tachistoscope **41**
'take the best' heuristic 360–1
Talairach coordinate system **454**
tangential dipoles **436**
target 49, **174**
task-switching 280
teleological explanations 13
template **454**
template matching **105–6**
template theory 325
temporal filtering **456**
temporal grouping effect 289–90
temporal lobe epilepsy 472
temporal lobe injury 472
temporal order judgement task 59
temporal resolution **431–2**
termination 84
texting skills 629–31
texture gradient **75–8**
thalamus 540–1
thematic role assignment **189**
theory of mind deficits 481
theory-theory of concepts 155, 157–8
thiamine deficiency 237
thought experiments 150–1, 159, 597
threat, attentional bias 531–2
3D object-centred description 82, **108**
3D object recognition 108–18
threshold value **607**
time-frequency plot 438
top-down processing 9, **70**, 95, 177
topography **437**
TOTE units **14–15**
touch 104–5
towards sentences **203–4**
Tower of Hanoi 309
Tower of London 310–11
TRACE model 175
training **608**

trait emotion **524–5**
transcranial magnetic stimulation (TMS) 93, 431, 432, 461–2, 465
transfer appropriate processing (TAP) 255–6, **244–5**
transitivity **339–40**
transmission of activation 640
transparency 408–10, 663
transposition error 289
traumatic brain injury 488–90
traumatic events 565–6
trials **437**
truth table 368–9
Turing test 15, 16, **597**
2½D sketch 82, **85–6**
two-deficit account 483–4
two-dimensional object recognition 105–7
Type I/II processing (memory) **231**
Type I/II processing (reasoning) 389, 390
typicality effects 147–9, 153–4
typicality ratings **147**

U

understanding **614**
unequal variance model 248
uniqueness point **174**
units **607**, 643
universality 663
unobservable 10–11
up-regulation **541**
utilization behaviour 281

V

V1 90
valence **518**
validating search procedure 378
values **620**
variable **613**
variable physical realization **671**
vascular dementia 475
vascular depression 487
vectors 85
vegetative state 598
ventral stream 57, **91–3**
ventral visual-limbic pathway 128
ventriloquism effect **55**
verbal analysis 304
viewer-centred description **85**, **108**, 122
visual attention 40–8
visual hallucinations 480
visual illusions **65**, 66–7, 408
visual neglect **65**
visual perception
  approaches to 69–70
  constructivist approach **87–90**
  eye 68
  Gestalt movement **71–2**
  Gibson's theory 73–81
  Marr's theory 81–7
visual system 90–6, 410
visual word recognition 176–82
visuo-spatial sketchpad 273
visuo-spatial working memory 273–4
vocabulary acquisition 285–7
von Neumann machines 15

W

Wada test 421–2, 471
Wason's selection task 379–86
water jar problem 307, 311
wavelength **35**
weak artificial intelligence 581
weight **607**, 643
Weiner, Norbert 14
well-defined categories **163**
Wernicke's aphasia **405**
'what' system 91
'where' system 91
winner-takes-all 175–6
within-category discrimination **102**, 117
word fragment completion 253
word-initial cohort **173–4**
word-length effect 274–5
word meanings 184–7, 200
word recognition
  spoken words 170–6, 179–80
  visual words 176–82
word recognition units 125
word stem completion task 583–4
word superiority effect 177
words and rules model **644**, 647
wordspotting task **172**, 176
working backwards **322**
working forwards **322**
working memory **266–95**, **578**
  consciousness 600–1
  executive processes 279–85
  long-term 325–6
  multi-component model 272–4
  phonological 274–9
  span 271, 280–1
  structure 272–85
  visuo-spatial 273–4
working self **548**, 558–60
written language 198–207
wug test 183

X

x-ray problem 306–7, 317

Y

Yerkes-Dodson law **521**
Yonelinas's theory 248–50
Young's dairy study 121–2

Z

Zajonc, Robert 539–40
zombies 596–7